Grounding Geographic Information in Perceptual Operations

Simon Scheider

Westfälische Wilhelms-Universität Münster, Germany

IOS Press

Abstract

In this thesis, I propose a method of constructing semantic reference systems for geographic information based on reproducible observations. Similar to spatial reference systems, semantic reference systems consist of formal theories (*reference theories*) with conventionally established interpretations into *perceptual and constructive operations* (such as pointing to a physical monument, and describing locations relative to it). They can be used to annotate data and describe and compare their semantics.

The thesis addresses the grounding part of reference systems. I argue that persisting problems of information ontologies, namely the *grounding problem* and the *problem of reference*, could be solved using certain principles of construction and imitation in terms of perceptual operations. These operations focus joint human attention on pre-conceptual cognitive mechanisms, i.e. 'Gestalts', in the perceived space around the body. They also allow an observer to relate foci of attention based on the Gestalt. The memorized relations are expressed by observation predicates, to be established by convention in terms of speech acts of an observation language. These speech acts can be shared, and thus represent lowest level information items. I propose a kind of "practical constructivism" guided by a formal language. The idea is to describe data categories in terms of observation predicates, i.e. "bottom-up", in order to reconstruct the underlying observation procedure, instead of presuming abstract concepts.

For example, bodies, surfaces, and different kinds of media in the human environment are grounded in terms of perceived affordances. Object properties such as waterdepth are grounded in experiential geometry, visual surfaces and media of diving. I propose a corresponding reference theory and put it to a practical test in this thesis: I define an essential road network category, namely a junction, and test the definition in Open Street Map.

I am sure that in every case the exact degree of opacity of reference will be made entirely transparent, even though the roots of reference must be an even dirtier subject than reference itself, which as we all know is dirty enough.

Nelson Goodman
in the introduction of Quine's *The Roots of Reference* [137]

Acknowledgments

There are quite a few people without whom this thesis could not have been written. There is, first of all, the inspiring input of the whole MUSIL research group at Münster I have to thank for. I especially would like to thank Prof. Werner Kuhn, who raised my interest in semantic reference systems and obviously considered my first (far-off) draft promising enough to give it try. I am very much indebted to my co-authors Krzysztof Janowicz, Florian Probst, Carsten Keßler, Jens Ortmann, Anusuria Devaraju, Johannes Trame, Jörg Possin and Tomi Kauppinen. I especially have to thank Patrick Maué, who offered me the opportunity of contributing to the Envision project, and the DFG, who supported most of the work by funding the Semantic Reference Systems project.

Furthermore, I owe a lot to my former colleagues at the Fraunhofer Institut für intelligente Analyse und Informationssysteme, Sankt Augustin. I would like to thank the whole KD team, especially Hans and Angi Voss (Angi for her patience in teaching the basics of formal specification), Natasha and Gennady Andrienko, Daniel Schulz, Michael May, Dirk Hecker, Roberto Rösler, Axel Poigné, Stefan Wrobel and Christine Körner. They all helped shaping early ideas in one or the other way.

Last but not least, I would like to thank my friend Jens Schütz for lending me an ear every now and then, and especially my wife Susi for her support.

Contents

List of Figures

List of Tables

Acronyms

AI Artificial Intelligence
FOL First-order Logic
GDF Geographic Data Files
ISR Intermediary Spatial Reconstruction
OSM Open Street Map
OWL Web Ontology Language
POI Points of Interest
RDF Resource Description Framework

Chapter 1

Introduction

Eine Lösung des sogenannten Welträtsels wird es nie geben,
weil das meiste, was uns rätselhaft erscheint,
von uns selbst geschaffene Widersprüche sind,
die aus der spielenden Beschäftigung
mit den blossen Formen und Schalen der Erkenntnis entstehen[1].

— H. Vaihinger [197]

What is the *intended interpretation* of geospatial data in terms of reproducible experiences? How should places on a digital globe be interpreted on the earth surface? And how can their spatial relations be reconstructed? How should road network databases be interpreted in terms of observable traffic infrastructure? And how is data about waterways and their depths to be interpreted in terms of observable water bodies? These are among the concrete questions whose general answer is sought in this thesis.

I propose a *meta-theory of information grounding*. It may help to solve some of the puzzles that occur when trying to describe the semantics of data. In particular, it is intended as a formally explicit way of describing georeferenced data, whether originating from volunteered or technical sources.

Why should one decide to describe the semantics of data? Data is usually generated with a concrete and understandable purpose, so do we have a problem at all? As discussed in Section 1.1, one is struck by the problem most directly in the role of a data consumer with an own application purpose who is not able to directly communicate with the producer. This is often the case when consuming sophisticated data on the web, such as *geospatial data services* [82], or when consuming data from the *sensor web* [170]. But it is also apparent in *classical data trading*.

[1] "A solution of the so-called world-riddle will never be found, since most of what appears to be puzzling are contradictions created by ourselves, originating from playfull preoccupation with the mere forms and shells of understanding."

For example, one may want to use *volunteered geodata* such as *Open Street Map* (OSM)[2] in order to build a navigation service. The correctness of a routing result thereby depends on the availability of correct observations, e.g., whether turn restrictions were mapped at junctions [129]. But considering the current OSM vocabulary, one remains unsure where the necessary information for navigation are hidden, and whether it is provided by OSM at all[3]. Also, OSM users have the opportunity to describe map features via tags. However, the tags that are used to describe points of interest (POI) often do not make clear what the *interest* in a specific point is. That is, they do not provide sufficient information about what is afforded by the POI: The tag `cafe` is used to describe *coffee shops* in New York as well as *Kaffeehäuser* in Vienna. If a user wants to have a beer, a place tagged `cafe` in Vienna is perfectly suitable, whereas a coffee shop tagged `cafe` in New York is not.

The problem also occurs with *sensor data*, for example if measurements are only described by SI units[4]. A speed value of 2 given in meters per second just tells us that there is something moving, but we cannot even tell whether it is a car on the road, or gravel on a slope, water in a riverbed, or whether it is just the allowed speed on the road. In this thesis, I will call this the *problem of data grounding* [63]. It occurs when it is not clear what kind of observation certain information refers to. Dealing with the semantics of geoinformation in terms of observable properties (such as temperature, precipitation rate, or traversability of a road), we also face the related problem of finding an appropriate level of description (*abstraction problem*). For example, precipitation rate can be measured by a tipping-bucket or a standard rain gauge. However, the meaning of a value of five liters of rainfall in the last 24 hours is independent of the concrete form of the sensor, and thus should be described on a higher abstraction level.

The need for data semantics exists and grows with the web's increasing role as a large interlinked data hub [14]. But a closer look reveals that despite technological progress, the grounding and abstraction problems remain largely unsolved to the present day (Section 1.2). The reason seems less that there are no computable techniques available that allow one to describe, query, and retrieve the semantics of data [13]. It seems more rooted in an *unsatisfactory understanding of what it is that needs to be described*. Current approaches seem to be stuck in "playfull preoccupation" with technological artifacts, i.e., with the "mere forms and shells of understanding", which Hans Vaihinger described so aptly in the citation above.

[2]See `http://www.openstreetmap.org/`.

[3]This seems a serious obstacle for OSM to compete with GDF standard data, compare [129] and Chapter 7.

[4]Le Système international d'unités, see `http://www.bipm.org/en/si/`.

1.1. Semantic heterogeneity and semantic strategies

In the early days of database research, issues related to database semantics played a prominent role in 'conceptual database design'. This began to change at the beginning of the 90's, and nowadays those issues do not appear to be part of mainstream database research [16]. The problem of data semantics regained attention in research areas concerned with *context-poor communication environments*. In 'information system integration', it was recognized since the 90's that *semantic heterogeneity* is a major problem in realizing interoperability [169]. In order to handle arbitrary information requests, users and providers need to mediate their understandings. *Semantic heterogeneity* occurs if *semantic interpretations of user and provider are not compatible with each other*[5]. Because a special case of a context-poor communication environment is the web, the idea of the 'semantic web' was launched later in 2000 [13]. What are the main strategies to deal with semantic heterogeneity so far?

Natural language strategy. A common approach to the problem is known from librarians. It consists in describing data by taxonomic schemes[6]. But as Shirky argued [171], taxonomies vary according to the purpose of categorization and indexing, which is different, for example, in a library and in the web. A common approach in standardization communities was to describe data with semi-structured metadata [169]. This approach has the same problems. As I will argue in Section 2.2, all natural language strategies have the problem that they rely on a common communication context, which cannot be presupposed in context-poor environments.

Formalist and ontologist strategy. From the very beginning of research in information semantics, *ontologies*, formal theories of the world as we know it [65], were proposed as a major tool for meaning description. At the same time, the sufficiency of *formal symbol systems* to construct intelligent agents [119] was debated in artificial intelligence (AI) [164]. One argument was the *symbol grounding problem* [63], the problem that declarative semantics expressed in terms of formal symbols gives rise to an infinite regress (Section 1.2).

Since then, information ontologies have been successfully used for describing data. This can be seen by many successful applications of light weight ontologies, for example in order to structure a whole web of knowledge with Wikipedia [15], or

[5]In Chapter 2, I propose to conceive semantic heterogeneity in this way. The term becomes very problematic if understood in terms of a "non-matching semantics" in analogy to a non-matching syntax [169]. What exactly should be matched here? Even more problematic is to speak of information systems not "understanding" the semantics of a user's request [169]. Information systems do not understand anything, only people do.

[6]For example by structured hierarchical vocabularies like *WordNet*, see `http://wordnet.princeton.edu/`.

in order to query cultural artifacts on the web [37], but also by the relative success of top-level ontologies like DOLCE[7] or BFO[8]. But even though the available ontological tools have matured in balancing expressiveness and computability, fundamental debates about their adequacy for a semantic strategy never stopped [171]. Nicola Guarino has recently argued [57] that the reason for the slow success could be a missing *ontological level* above and beyond formal representation, which has built in ontological distinctions reflecting our way of understanding the world. I agree and will discuss the fitness of the pure formalist approach in this respect in Section 2.2.

Cognitivist and objectivist strategies. In the AI community, two opposing early ideas were to (1) base semantic theories on human cognitive abilities (common sense knowledge) and to (2) base them on objective reasoning and natural science [67]. This debate, originating in the philosophical opposition of nominalism vs. realism, recurred in linguistics [90], and has been continuing to the present day in the information ontology community [114]. Are human conditions of reason more an obstacle to or an indispensable condition of representing knowledge? From a practical perspective, it was recognized quite early in spatial reasoning that the main challenge to represent spatial concepts lies in the need for different and context dependent world views [45]. In Section 2.2, I will argue why an objectivist strategy has difficulties of coping with the problem of reference in general, and why cognitivism alone fails to make the point of inter-subjectivity.

Collaborative and pluralist strategy. One progression was towards more collaborative approaches based on user-generated 'folksonomies' [110]. The underlying idea is that once semantics is not prescribed, it can *emerge* in an interactive community. Volunteered geodata [53] are managed by following this guiding idea. The strategy has the advantage that user and producer intermingle or more or less coincide, and thus can share a context. As the limited quality of OSM data - documented by the debates inside the community[9] - show , however, the problem of semantic heterogeneity has not disappeared, but seems rather transferred to a new level. What is sometimes called 'the wisdom of the crowd' seems to be affected by a kind of *"topsy-turvy" semantics*: Individual understandings are still divided by semantic heterogeneity. But this heterogeneity remains undiscovered, because it becomes blurred on the level of the crowd.

Another progression was to use semantic similarity [73] and schema mapping tools [38, 200]. This approach accounts for semantic heterogeneity by acknowledging pluralism, and tries to do the best in building bridges between heterogeneous worlds. A new variant of this strategy is the linked data web [14]. It is not a bad

[7]Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [109].

[8]Basic Formal Ontology (BFO) [118]

[9]Compare for example `http://wiki.openstreetmap.org/wiki/Category:Quality_Assurance`.

compromise between worlds, but rather a proposal to collaboratively establish links between different worlds. Janowicz argued [71] for rediscovering the opportunities of space and time in organizing this pluralism, instead of arguing against it.

Meanwhile, the problem has remained pretty much the same: heterogeneous information communities need data from diverse external sources in order to produce new products and services, but these products were mostly *not intended by the data suppliers*. Therefore the discovery, retrieval, integration and query of such information requires from the user to be able to reconstruct its *intended meaning*.

1.2. The problem of grounding semantic descriptions of data

As Kuhn argued in [84], despite the notoriously difficult philosophical questions involved, *semantic interoperability* can be seen as an *engineering problem*, namely that of effectively constraining interpretations towards the ones that are considered allowable. Solutions should be based on minimal and sound methodological assumptions and a clarification of the scope with respect to this goal.

What is the precise problem we are tackling, what are existing approaches, and what would be the scope of its solution?

The *grounding problem of information science*, first formulated by Harnad [63], is the problem that referencing the meaning of primitive notions of a formal theory in another theory gives rise to infinite regress. The meaning of formal symbols in this case stays in the realm of symbols and thus never gets into any determinable contact with the world [164]. From a practical viewpoint, if this chain cannot be cut and tied to some definite reference in the world, then we are left without a means to *refer* to one and the same thing when communicating with each other. This is called the *problem of reference*, which is the practical manifestation of the grounding problem. The problem of reference appears in communication situations in which participants do not share a common observation context to which they can refer. A simple example is a measurement scale, which, although it can be described algebraically [182], is not definable and needs measurement units for its unambiguous establishment. These units can only be established in a context where communicators can mutually refer to reference phenomena. For example, the 'mètre des archives' used to be the reference object for a meter. The problem of identity and reference is also a known obstacle to the vision of the semantic web, and currently discussed in terms of 'same as' links [62].

One suggested way out of the regress cycle is based on referencing by inherent capabilities of the human body. Such an approach parallels the idea of embodied semantics, which was most prominently put forward by cognitive scientists like Lawrence Barsalou [6] and cognitive linguists like Leonard Talmy, Ronald

Langacker, Mark Johnson and George Lakoff. These authors have alluded to the grounding problem for the realm of human cognition in general, not only for language. According to them, abstract or metaphorical concepts in thought are closely linked to sensory-motor perception patterns. From the side of language cognition, there are strong arguments in favor of such a link, but it remains to show how it can be used for purposes of information grounding.

The grounding problem has also been recognized in robotics and AI and approached in a similar way [199]. Here, the main argument is that intelligent systems cannot evolve unless artificial reasoning is grounded in sensory-motor perception and interaction in a social community of robots [199, 176, 8]. But the focus of these researchers is on understanding the mechanism of language acquisition and evolution, not on semantics of human understandable languages. Therefore, Luc Steels' claim [177] that the grounding problem is "solved" seems misguided if we regard it in the context of information grounding. One progress that has been made is that many researchers agree on embodiment as a way of solving it. But a 'semiotic network' established by a group of robots [177] still runs into the grounding problem if an extraneous agent tries to understand what these robots are talking about. Note that this is an exact analogue of the situation we are facing in the semantic web: *The problem is not that machines are unable to communicate, but that humans misunderstand each other if communicating via machines.* Nevertheless, Steels' work makes clear that a grounding solution is a *collective achievement by social interaction in a context.* [177].

In the field of information science, for example in the geospatial web, the grounding problem has a particular relevance: Matching web services and data with the expectations of the user experiencing an environment is still a big challenge [82]. This is because formal metadata, e.g. ontologies, even though it constrains the data's use, still allows for interpretations or mappings that were not intended by the data providers [56]. It is therefore not surprising that *information ontologies as such are not a solution* to the problem, and that data integration and schema mapping are still among the key research topics of the latest research initiatives in the web, such as linked open data [14], the semantic sensor web [170], or big data [202].

The grounding problem has also been tackled in Geographic Information Science (GIScience): For example, by tying ontological terms to known datasets [11], which constrains these in potentially useful ways, but defers the grounding problem to the symbol system of the instance data. Frank proposed that objects and social reality can be considered as ontological tiers based upon observable reality [42]. Couclelis [28, 27] suggested a teleological information hierarchy, in which the information concept of *purpose* implies all others and is located on top, while lower level concepts are obtained by "thinning out" the ones on the upper level ('semantic contraction'). These latter two approaches are comparable to and useful for the one proposed in this thesis, since they regard information as a product of purposeful construction. But they remain to be spelled out in many

respects.

What is required is an *epistemological* view on experiential qualities [108], which is only starting to develop in the ontology community [107, 131]. There exist useful areas of research on qualities oriented towards the empirical structure of qualia and mental spaces, related to psychophysics and measurement theory. Suppes has extensively investigated the formal structure of perceived and measured spaces [180]. Adams and Raubal have proposed an algebra for conceptual spaces [1] as proposed by Gärdenfors [48]. Such formalisms are useful for describing domains of experience, but still fall prey to the problem of reference: what individual phenomena do the intrinsic dimensions of a conceptual space refer to? How can an observer reidentify a point of this space in his or her environment? Qualities exist on all abstraction levels and do not incorporate clear experiential references [159].

In this thesis, I propose that the wanted epistemology could be a *pragmatic* one. There is one particular pragmatic idea on which this thesis is based: *Semantic reference systems*, first proposed by Kuhn [81, 87]. This concept draws on the analogy to *spatial reference systems*. Spatial reference systems are grounded mathematical ellipsoids that allow to refer to locations on the earth surface. They are remarkable technical inventions: they not only provide a computable form of a location, but also allow geodesists to *reidentify individual locations* in their environment. That is, they solve the grounding problem for loci by establishing physical references for "anchoring" an ellipsoid with respect to the earth surface. These reference phenomena are called *geodetic datums*, and consist in conventional standard directions and positions that geodesists can refer to in their observable environment[10]. Semantic reference systems, in analogy, are formal theories anchored in conventionally established observation procedures. The formal framework may be analogous to a conceptual space, while it is the reference observations that provide the grounding.

A semantic reference system consists of two parts, the *semantic reference frame*, i.e., a formal theory such as an ontology, and a *semantic datum*, which provides its grounding [87]. The reference frame can be used to describe or define a notion. A semantic datum coordinates interpretations of the primitives of a reference frame among interpreters. As all nonprimitive symbols in a theory are definable from the primitive ones, a semantic datum can fix one particular interpretation. To this end, it must be based on some *method of grounding*.

One existing approach to semantic reference systems is the one of Probst [131, 130], who revealed a necessary human conceptual basis in terms of *reference spaces* for qualities. But reference spaces are still based on the presumption of quality spaces, and not on the actual performance of observations. Using

[10] A geodetic datum for the positions on a Bessel ellipsoid consists for example of a named spot on the earth's surface like 'Rauenberg' near Berlin (Potsdam Datum), and a standard position and orientation for the ellipsoid.

metaphorical concepts, such as image schemas [83], has a similar problem in terms of reference that will be discussed again in Section 3.3.1. A solution has to ground the ontology of qualities and quality spaces in *reproducible observation procedures*.

Another approach tries to account for qualities by generalizing procedures of measurement and observation [85, 175] in the manner of a *sensor science*. This can be particularly useful for the semantic sensor web [170], as it provides abstract descriptions for observation processes. Janowicz [72] has recently argued for *observation-driven ontology engineering*.

From an information grounding perspective, it remains an open question what concrete observation procedures we as observers have at our disposal, in how far they can be reproduced and inter-subjectively referred to by us, and how they can be used in the construction of qualities. I propose to call a theory *grounded*, if there exists a convention for fixing the interpretation of its primitives into repeatable and inter-subjectively available observation procedures. How this can be done, and what operations are at our disposal, is the main subject of this thesis.

1.3. Scope of the thesis

How can data-providers indicate to the users the intended interpretation of their data set? In particular, how can they *refer* to the underlying *operations and observation procedures*? These questions were my main motivation to write this thesis.

From a theoretical perspective of semantic engineering, the thesis addresses the corresponding *problem of reference* for geographic information, i.e., how to refer to the things we mean by the terms used in data sets. More precisely, it aims at identifying *concrete reference mechanisms* that allow to *establish a particular interpretation* of a data set. Through analyzing these mechanisms, we can understand how reference systems are actually grounded, and in the long term, we may arrive at a general method of constructing them.

Thereby, I start from the following *hypotheses*:

1. The generation of semantic reference is based on a small set of operations.
2. These operations are used by an observer to reconstruct and refer to a large variety of ontological perspectives.
3. There is a basic reconstruction (called the *meaningful environment*) that allows to refer to the immediate human environment in the sense of Gibson [50].
4. This reconstruction accounts for the meaning of central terms and categories in geodata.

The research challenge consists in making justified suggestions for these reference operations, in proposing an unambiguous language about them that helps

in establishing reference, and in showing that concepts of the meaningful environment can be reconstructed with this language, such that they comply with the meaning of essential terms in geographic information.

I address this challenge in terms of the following concrete *research questions*:

1. Which kinds of reference operations can be justified from cognitive research, and how to establish a precise language about them?
2. How can geometrical and topological concepts, environmental media and qualities of the meaningful environment be reconstructed by an observer, starting from a small set of primitive operations?
3. How can road network categories like junctions be reconstructed in this meaningful environment?

The proposed *solution*, which will be developed in the following chapters, can be summarized as follows: Based on our attentional apparatus and shared perceptual Gestalt mechanisms, we (as humans) can establish an observation language by learning certain acts of predication and referencing. The thesis proposes a range of logical primitives for such a language based on perceptual operations, and demonstrates how they can be used to reconstruct major geodata concepts.

This is the core of an effort to provide a meta-theory for grounding digital information. The meta-theory can be used to generate particular reference systems for a semantic domain. A semantic reference system serves to ground data descriptions or annotations, such as OSM tags or RDF concepts in the semantic web. In the following chapters, I will develop a corresponding meta-theory.

In Chapter 2, I will discuss in how far current semantic strategies, which were discussed above in Section 1.1, fail to solve the grounding problem.

In Chapter 3, I will propose an *operational approach* to the problem of language reference, which is, as we already saw, the heart of the grounding problem. I argue for an attentional apparatus that allows humans to perform *perceptual operations*, namely *predication and referencing*. These enable us to join our attention on shared perceptual Gestalt phenomena in our space around the body, and thereby to solve the problem of reference.

I will then suggest *constructive means of data grounding* in terms of a technical observation language (Chapter 4). This language needs to be established by convention and introduces formal observation predicates as primitives. These denote results of perceptual operations stored in memory of an observer. The language also accounts for the freedom of reification, i.e. of constructing abstract entities. The most relevant abstraction tool used in this thesis is the reification of *wholes*. These are maximal classes of foci of attention self-connected by some perceptual relation.

In Chapter 5, I will discuss *perceptual sources of data grounding*, including justifiable suggestions for observation predicates. These predicates account for different domains of experience of a human observer. They are justified by empirical evidence of underlying perceptual operations.

In Chapter 6, I will demonstrate how these sources can be fruitfully combined in order to reconstruct major domains of experience underlying geographic information, including experiential geometry, length, media, substances and surfaces. These can be used to describe meaningful things like waterdepth, media of locomotion and places.

In Chapter 7, I will apply the theory to the domain of road networks and OSM, before concluding in Chapter 8.

Part I

The observational roots of data

Chapter 2

The necessity for grounding data

Consider, for example, the case of a geopolitical globe as a model of the Earth. The (informal) questions to be asked have to do with the existence and (relative) position of features on the Earth's surface, [...]. The questions about the model are answered by direct observation of the model by a human, aided perhaps by a string/ruler/compass [...]. The mapping of questions and answers is based on the scale reduction of the Earth's spherical surface to that of the globe.

— Borgida and Mylopoulos [16]

> "A data source is useful because it models some part of the real world, its subject (or application, or domain of discourse). The problem of data semantics is establishing and maintaining *the correspondence* between a data source and its intended subject" [16]. (my emphasis)

Probably, few readers will disagree with this statement. But the problem with it - as with many such agreeable statements - is that it is almost as flexible as a rubber band. It is sufficiently ambiguous that every reader can almost arbitrarily interpret his or her own view on the matter. Thus it causes more confusion than understanding. In the sense of ontology engineering as a way of effectively constraining the allowable interpretations [84], it would be necessary to work out what is meant with the notions *real world* and *correspondence*. I suggest not to do this.

Instead, it may be more successful to ask how human interpretation can be effectively constrained such that the problem of reference is practically solved. In this chapter, I will argue in how far current semantic strategies are challenged by this problem[1]. I start with a scenario taken from the former quotation above, which leads us to some important insights.

[1]This chapter is an abbreviated version of ideas previously presented in [151].

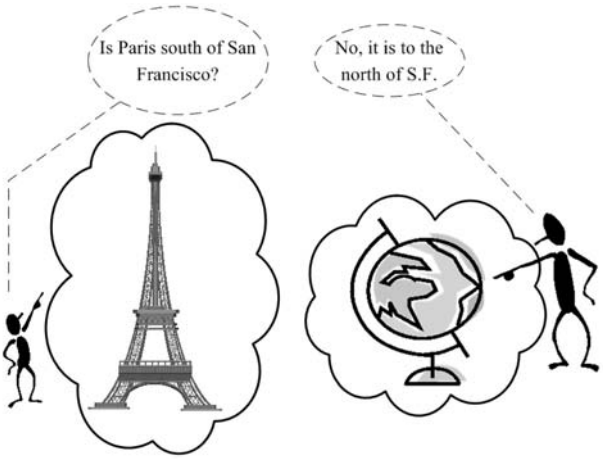


Figure 2.1: A database query model.

2.1. Data queries need to relate human interpretations

Borgida's and Mylopoulos' revised idea about data semantics [16] is helpful to see the problem more clearly. They propose to see *data as a 'model' whose purpose is to answer human questions* about a subject. They illustrate this by the example of a data query about locations on a globe (compare first quotation above). There are two far-reaching lessons hiding behind this simple scenario:

Data needs to be interpretable by humans: First, this case suggests that a data set should *have meaning in the same sense as natural language has*. Otherwise the data query could not produce meaningful answers in natural language. This notion of meaning essentially involves that the data set should be *interpretable in terms of human experience and human thought*. The computation of the relative location of Paris and San Francisco (Fig. 2.1) requires this. It seems to be a rather trivial fact, but it is not: main stream computer science since the times of Ted Codd's invention of the relational database model has sticked to another view. Semantics was factored out during run time of a database: Database semantics was managed by the operational environment of a database system, i.e. its database administrator, and "If you wanted to know what the data really meant, you'd have to talk to the administrator"[16]. Database queries — following the 'physical symbol systems hypothesis' [120] — were considered to be formal manipulation of symbols without any necessary further meaning. But it is this very precondition of *human interpretability of databases*, which seems to make the implementation of an intelligent digital program, like the question-answering *'Turing Machine'*, a mere fiction[2].

[2]John Searle argues in [164] convincingly that symbol manipulation is not *sufficient* for

Interpretations need to be coordinated: Second, the mapping between questions and answers in the data query is based on a *relation between different interpretations*: the relation of the earth surface to the globe model. Note that what is being related here are observations. The person posing a question must be able to interpret an answer in terms of a place on the observed earth surface. But the answer was derived by another person from a separate observation of the globe surface. *This relation is exactly how — through a chain of interpretations — human-interpretable questions are linked to human-interpretable answers.* And it is the reason why the database query actually works, that is, why it is able to deliver meaningful answers to humans.

What exactly is meant here with the notion of *'interpretation'*? Something very similar to von Glasersfeld's concept in [51]: An interpretation is a *decoding of symbols in terms of thoughts and experiences.* More specifically, it is a *mind specific activity of a human interpreter S* taking an *immediately experienced object X (the semiotic object, e.g. a symbol)* and producing *a not immediately experienced result Y*, which is not part of X. X could for example be the question that person S tries to answer by interpreting it on the experienced globe Y. If X — like in our case — was created by another person A, the 'author', then A's own individual interpretation similarly may have produced M, the *'intended meaning'* of X, which is the earth surface experienced by person A (see Fig. 2.2).

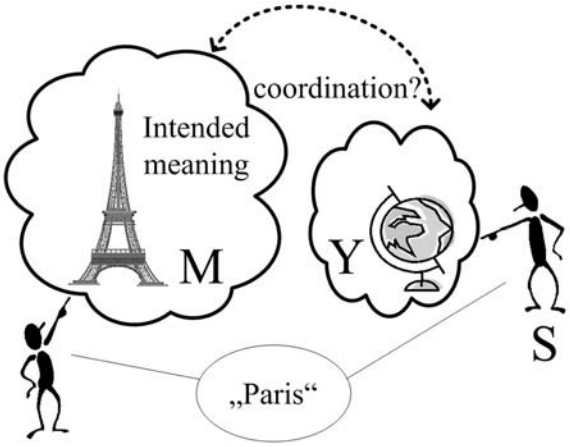


Figure 2.2: How to coordinate interpretations is the real challenge for a successful database query.

The view that the semantics of a *formal vocabulary* requires interpretation in terms of thoughts and experiences (sometimes called 'conceptualization' in order

understanding a natural language text, because *understanding* involves the possession of intentionality: the possession of mental states, like e.g. beliefs, desires and intentions, that are directed at states of affairs in the world.

to stress their interdependence), has been common from the early days of logic based research in artificial intelligence[3]. This view seems to be agreeable among ontologists (see e.g. Guarino [56]), because it contradicts few philosophical world views (apart from, perhaps, a platonistic one).

But I would like to stress two aspects of potential disagreement. First, thoughts and their interpretative association are *artifacts of the individual human mind* [51]. The two interpretation activities in our example produce results in two different heads, so *Y never will be the same as the intended meaning M*. This fact about interpretation therefore stands in sharp contrast to the 'conduit metaphor' of language [144], or any naive idea of a 'universal concept'. Second, the *semantic rules* seem to be a private affair in the first place. If we strictly define a symbol in terms of thoughts or experiences, we arrive at a close association of thoughts. If we loosely relate a symbol to our web of beliefs, there can be much more variation in terms of facts. These degrees of freedom will be my main concern in this chapter.

Semantic interpretation poses a challenge to every database query. However, as we see in our example, *the private interpretations are often successfully coordinated in such a way as to achieve meaningful answers to questions*. I suggest that this is so because *questions and answers are both grounded in common operations*: the answer of S is grounded in geodetic measurements, which are themselves grounded in the human observation of the environment, and this indirectly observed environment of S happens to be closely coordinated, *on this level*, with the directly observed environment of A. The two interpretations are on a certain level on which the bodily operations — for example the measuring of distances on the globe by S and on the earth surface by A — are *mutually referable* among persons S and A, and therefore related.

This is the implicit idea contained in Borgida's and Mylopoulos' example:

> Successful database queries need to relate questions to answers by coordinating their human interpretations.

Once this problem of relating interpretations is solved, the problem of semantic heterogeneity is solved, too, as any query posed to a data set could be given an answer related to the question's *intended meaning*. But the former turns out to be the real problem, because language interpretation is a largely indeterminate process, and coordinating this processes is a challenging task.

[3]"In making our definition [of semantics], we assume the perspective of the observer [...]. We have a set of sentences and a conceptualization of the world, and we associate the symbols used in the sentences with the objects, functions, and relations in our conceptualization" [49, Chapter 2.3].

2.2. The indeterminacy of language interpretation

In this section I will discuss evidence and arguments for the view that semantic interpretation of language in general is a vague, underconstrained and indeterminate process. The indeterminacy of this process challenges the *existing approaches to semantic engineering* (see Section 1.1), because it undermines their assumptions: Determinism of scientific thought in natural science, established usage of natural language, and precision in declarative semantics of formal languages.

2.2.1. The argument of indeterminacy of empirical theories

The first argument is one of epistemology, that is the justification of knowledge. If a web of beliefs[4] was shared among interpreters, the coordination task could be reduced to *negotiating a common vocabulary* for the notions involved in those beliefs. The argument has much to do with the methodological implications of *philosophical realism* (also called *metaphysical objectivism*) [20].

If we assume that our web of beliefs can be harmonized by approximating the 'real world' through observation and scientific reasoning, then this could be a way of coordinating interpretations. If the real world consists of discrete objects and their relations, then agreement on interpretation is just a function of the correctness and completeness of representing those relations in thought and language. Perhaps one could admit that only experts really have access to these correct descriptions, like Putnam did [133], and thus require ontologies to be constructed solely by experts.

Realism is in fact quite often used as a justification for a certain methodology of ontology engineering. One example is Barry Smith's proposal [173] to replace the notion of a *concept* as the subject matter of ontologies by "the universals and particulars which exist in reality". The idea is that the ambiguity of ontological terms as well as the existence of different ontological views on a subject matter, which seem deeply entangled with the imperfectness of human perception, will disappear once ontology engineering is "devoted precisely to the representation of entities as they exist in reality" [173], and not to mere linguistic or cognitive artifacts. In this view, people *hold different views on reality due to human misconception*, but once they strive for a better (less subjective and more natural-science compatible) view, their conceptions will converge. A comparable philosophical view stands behind many attempts to formalize, once and for all, a so called *universal ontology*, in which agreed human knowledge is encoded.

It is well known that scientific realism is confronted with a large list of pow-

[4]In epistemic logic, knowledge is considered as true belief, i.e. the special case of believing in true propositions. But knowledge is a much debated notion, and truth even more. In this thesis I am talking about webs of human beliefs in that general sense, i.e. about interrelated proposition that are *accepted* as being true, regardless of being also called 'knowledge'.

erful criticisms[5]. I will not engage in any metaphysical discussions about realism and think that a 'basic realism', e.g. in Lakoff's sense [90], is beyond question. But I would like to discuss the *practical methodological problems* that one encounters when following a realist approach, in order to illustrate the effects of the indeterminacy of language interpretation.

Empirical Underdetermination. One of the most powerful arguments against philosophical realism is based on the so called *empirical underdetermination hypothesis* [89]:

> For every empirical theory T there is an *empirically equivalent* theory T' which is incompatible with T.

The term 'empirical equivalence' can be explained in the following way:

> "Call two theories empirically equivalent just in case exactly the same conclusions about observable phenomena can be deduced from each."[20]

The argument of empirical underdetermination does not force one to assume that there is no objective world outside of human thought, or that progress in scientific reasoning is not possible. But it leaves no reason to think that sophisticated conceptualizations of the real world will *automatically and asymptotically approach a unique state of agreement*.

The power of the underdetermination argument can best be illustrated by an example. W.V.O. Quine [142] gave one: According to physical theory T, we live in a universe that can be described by an infinite 3-dimensional Euclidean space. However, according to theory T', we live inside of a 3-dimensional ball, and the more an object approaches its surface, the smaller it gets. Empirically, by using observation or measurement, it is impossible to distinguish between the two worlds that are described by T and T', because the measurement units will shrink in exactly the same way as the objects do. Nevertheless, the theories are incompatible because in T', an unobservable center of the universe exists, but not in T.

The idea of empirical underdetermination is a radicalization of the famous *Duhem-Quine thesis*. Empirical contradiction (often called *falsification*) can — according to Duhem [32] — never be accomplished for an isolated hypothesis. The argument is that *observable implications* never exist for a *single hypothesis*, but only for a conjunction with auxiliary premises. Falsification therefore only implies that any assumption, including the hypothesis, could be false, but which one is left undecided. The argument can be extended to *whole theories* as sets of sentences, their inference rules and derivable facts. As Quine put it: "Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system"[6]. He devises an alternative conceptual view on empirical knowledge, which might be called the *'undetermined fabric of knowledge'*:

[5]See [20, 19] for an extensive discussion.

[6]See [136]. Quine's critique is directed towards what he calls *reductionistic verification* in

> "The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even pure mathematics and logic, is a *man-made fabric* which *impinges on experience only along the edges*. A conflict with experience at the periphery occasions readjustments in the interior of the field.[...] Re-evaluation of some statements entails re-evaluation of other, because of their logical interconnections — the logical laws being in turn simply certain further statement of the system [...]. But the total field is so undetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to re-evaluate in the light of any single contrary experience." ([136], emphasis mine)

This so called *holistic* argument is not really in danger by saying that in concrete cases, researchers usually can agree on which parts of a theory to retain by referring to preferential rules and heuristics. The problem has practical analogues in the field of *knowledge discovery* and *statistics*: In order to learn a curve from observations e.g., it is always possible to choose among a set of linear and non-linear regression rules. But this choice is often not clearly decidable from empirical evidence: it needs a *theoretic bias*. There are lots of heuristic strategies to select from [116], e.g. simplicity and *Ockham's razor*. But it is not the case that this rule — or any other theoretic bias — is most likely to be successful a-priori [174].

Taking these arguments into consideration, it seems naive to assume that even in the most accurate natural sciences, e.g. physics, there will emerge one *objective ontology* which contains all abstract concepts of the discipline. This of course does not mean that it is impossible to build a single meta-theory in physics. It just means that given the empirical facts, different ontological views must always be expected. The examples of *matter*[7], *mass-energy equivalence* [39] and the *wave-particle dualism* ([21], Chap. 9) may serve as arguments in favor of this view. But also the confusion about biological taxonomies ([90], Chap. 12) or the case of the duck billed platypus, as described by Eco ([35], Chap. 4), illustrate the case.

Observability. It has often been brought up against the empirical underdetermination argument, that it draws on the distinction between *'observable' and 'unobservable'* sentences in a theory, and that this distinction cannot sharply be made, or that it is theory dependent itself. First of all, observability seems to be so crucial a notion that I do not see how to do without it even in a realist setting. Furthermore, as Boyd [19] demonstrates, such a distinction must not be sharp, but could be a continuous transition from observable to unobservable, and can be sharpened if needed.

In fact, the notion of observability can be made precise in the way Quine did,

---

empiricist thought (popularized e.g. by Locke, Hume as well as Carnap), namely the view that every statement of a theory is translatable into a statement about immediate experience and thus can be evaluated in isolation.

[7] "It is fair to say that in contemporary physics, there is no broad consensus as to an exact definition of matter"[203].

for example in [140] (I will return to this in Section 4.1.3): Quine's argument is that natural language sentences vary in their semantic indeterminacies. There are certain *occasion sentences*, utterable only on the occasion, with relatively low indeterminacy and high observability, like 'it's raining' or 'it's a rabbit'. These sentences are called *observation sentences*: "An observation sentence is an occasion sentence that the speaker will *consistently consent to* when his sensory receptors are stimulated in certain ways [...]" ([140], emphasis mine).

The correct usage of observation terms is inculcated in the individual child of a language community by social training on the occasion, that is by the child's disposition to respond observably to socially observable situations, and the adults disposition to reward or punish its utterances ([143], Chaps. 1 and 3). Observation sentences are *the entrance gate to language*, because they can be easily learned directly by ostension *without reference to memory or theory* ([137], §11). In this way, observation terms actually spread far beyond the concretely observable situation and consistently recur in different theories [140][8].

Because the human 'fabric' of empirical knowledge is largely underconstrained by observation, different and incompatible conceptualizations of reality have to be expected at any time. This means that the semantic engineer confronted with semantic heterogeneity problems basically is left without the 'God's eye view', but not without *reference to collective experience*, and not without 'truth' or 'rationality', as Putnam argues[9].

2.2.2. The argument of indeterminacy of natural language use

One could argue that, because natural language is practically successful in conversations, in a way that seems to perfectly constrain the intended meanings of words and sentences, why not solve the problem of semantic heterogeneity by sticking to natural language descriptions? There are definitely parts of natural languages whose interpretation is strongly constrained by the language community. As John Searle [166] points out convincingly, language is a constitutive part of *constructed social reality*, especially in order to establish *'objective' institutional facts*, like e.g. the supreme court making a decision, or the assignment of a status like 'money' to a piece of paper. It is important to notice that there are *subsets of natural language which already come with coordinated interpretations*, because they are results of the collaborative construction of social reality. I would count Searle's *institutional facts*, but also Quine's *observation sentences* here, which were discussed in the last subsection.

[8]Quine also points out that in scientific discourse, observation sentences are the 'common ground' to fall back on [137].

[9]"'Truth', in an internalist view, is some sort of idealized rational acceptability, some sort of ideal coherence of our beliefs with each other and with our experiences, as those experiences are themselves represented in our belief system — and not correspondence with mind-independent or discourse-independent 'states of affairs'" ([134], Chap. 3, page 49-50).

One may think that a determined use and interpretation of language can be a social fact itself. But is the interpretation of an arbitrary word fixed in the same way as the interpretation of a certain piece of paper as money? It turns out that an important source of indeterminacy of natural language are its degrees of freedom in usage. Linguists have frequently pointed out that natural language usage must be a *creative and open process*, which makes use of these degrees of freedom to account for the fact that an unlimited variety of meanings must be expressible using only a very limited human lexicon [68]. One observable effect of this is the frequent *re-use or re-interpretation* of lexical symbols, as well as their *metaphorical use*.

Linguists have discovered many sources for semantic indeterminacy of language, e.g. *graded structures* and *prototype effects* of category words (compare the discussion in Chaps. 4, 5 and 6 of Lakoff's book [90]), like in the case of the word 'mother'[10]. Sometimes categories are used *metaphorically*, and it is unforeseeable which aspects of a prototypical meaning motivate those category variations under the umbrella of a single word, the border case of which are totally 'unrelated homonyms' (see Lakoff's example of the Japanese 'hon', Chap. 6 in [90]). Lakoff also observed (see [90] Chap. 8), that cognitive models underlying the semantics of language sometimes *need to be inconsistent*, because language can be used to talk about itself in the same sentence. For example, *negation* can mean to deny a fact but also a whole cognitive model.

It is this *context dependence* of meaning in natural language that makes it flexible as a tool for human communication, but does not render it appropriate for constraining semantic interpretations in a context-free setting. This is what is meant by saying that human language is 'imprecise', whereas formal logic is 'precise'. Negation in logic has one and only one interpretation, because its semantic rules are explicit. So it seems that formal languages are *necessary* to restrict interpretations effectively in context-free communication environments. But does this mean that formal theories are also *sufficient* for coordinating interpretations?

2.2.3. The argument of unintended semantic domains

One might think that ambiguity of language interpretation may be a problem only for natural languages. But as I will show in this and the next subsection, the problem exists for *natural* as well as for *formal languages*. Therefore semantic heterogeneity problems can always be found in both, and, as a consequence, *formalisms alone turn out to be an inadequate means to solve these problems*. It becomes apparent that the symbol grounding problem is a very essential constraint for all languages, which implies deep practical problems, because *meaning*

[10]The genetic, the nurturance, the marital and the genealogical aspects of motherhood are all fulfilled by a prototypical mother, but there are non-prototypical usages of the word, e.g. 'genetic' mother, stepmother and so forth, that have nothing in common, but seem to agree only with arbitrarily selected aspects of this list.

— *understood in terms of interpretation of symbols by humans* — is not conveyable in any language alone.

The first argument is concerned with the *identification* of what is called a *domain of interpretation* in model theoretic semantics. In first-order logic (FOL), a *signature* is a set of constant, predicate, and function symbols. Together with the syntactic rules of FOL it gives rise to a language (the set of well-formed formulas, or sentences). Now, in model theoretic semantics, to interpret a signature, sometimes called *a structure of the signature*, means

1. to identify a *concrete set (called 'domain', e.g. D)*, and
2. to associate each *constant* symbol with an *element of D* and each *function/predicate symbol* with a *concrete n-ary relation in D*.

Tarski's model theoretic truth definition tells us when a given sentence is true in this structure, in which case we can talk of the structure as a *model of the sentence*. Model theory is almost exclusively about this second aspect of a structure, that is it is assumed that D together with its concrete relations is given, and we just look at those interpretations that preserve the asserted truth of, i.e. *satisfy*, certain sets of sentences, called theories.

But what exactly does it mean that a *concrete set 'is given'*? The problem is that *meaning*, as I conceive it here, is an interpretation of symbols into a *specific domain*, namely into the domain of our thoughts, experiences and mental operations. As we saw, this domain is in no way *'given'* in the sense that everyone has equal access to it.

There is a method of making us aware of the personal mental domains we use when we interpret signs. It makes use of the idea of *analogical representations*, which are being discussed in the seminal works of Sloman [172] for artificial intelligence and Palmer [125] for cognitive science. Sloman argues for the existence of analogical representations including truth-values and even valid inference procedures: "Discovering the truth-value requires the application of semantic interpretation procedures in investigating the world" [172]. The idea is that if we investigate the world around us *and* if we interpret a visual sign, we *experience concrete relations* between parts of (what Sloman calls) a 'configuration', and these relations can be used to make valid statements and even inferences. For example, look at Fig. 2.3: It is a visual configuration that represents 4 objects ordered with respect to their tallness. We are immediately able to recognize an order relation among the 4 objects: 'Taller than' is a fundamental mental operation we are used to apply to spatial objects. Note that this operation has logical properties, e.g. if a is taller than b and b is taller than c, a is also taller than c (transitivity). Figure 2.4 is a representation of the tallness relation by another ordering operation, 'length'. Note that this representation does not have widths anymore. It is *poorer*, but also *less ambiguous*. In Fig. 2.5, 'Taller than' is represented by 'Points to'. Here, almost every similarity between representation and the thing represented ('iconicity') is lost, but an *explicit sign* for the relation

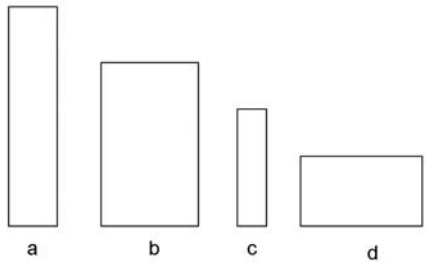


Figure 2.3: *Analogical* representation of 'Taller than'.

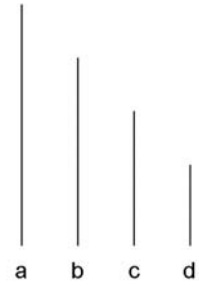


Figure 2.4: *Analogical* representation of 'Taller than' by 'Longer than'.

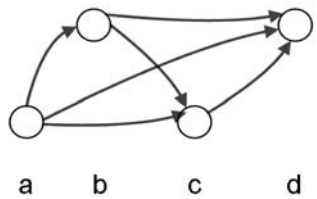


Figure 2.5: *Analogical* representation of 'Taller than' by 'Points to'.

$a \geq b$

$b \geq c$

$c \geq d$

$\forall x,y,z.\ (x \geq y \wedge y \geq z \rightarrow x \geq z)$

[function] Takes [arguments]

Figure 2.6: *Fregean* representation of 'Taller than'.

'Taller than' is present, the arrow. Figure 2.6 is actually a *formal first-order theory* representing tallness. There is only *one mental operation left*, namely *function application* (I call it 'Takes'), which has to be constructed while reading the text. Every semantically relevant mental aspect was explicitly converted to an atomic or constructible sign: the four objects and their immediate tallness-neighbors are written down into four facts. Other facts, e.g. that a is taller than d, can be deduced by applying the fifth fact, the transitivity axiom.

As we see, all of these representations need mental operations for interpretation, but to a different degree, which makes them more or less flexible to represent other mental operations. The most flexible one is of course 2.6, which Sloman calls a *Fregean representation*[172]. But the *power of constraining possible interpretations — better: of reducing the potential mental domains of interpretation — decreases dramatically from Figs. 2.3 and 2.4* to Fig. 2.6. There are two ways of applying a mental order operation in 2.3 and just one way in 2.4. Because length in Fig. 2.4 is iconic to tallness, Fig. 2.4 is even able to hint at the correct interpretation. But how many mental domains of ordering could be denoted by Fig. 2.6, *unless we already know* that it is supposed to represent only 'Tallness'? It could be interpreted in terms of every domain with a partial order, that is, *nearly our whole universe of thought*: natural numbers, real numbers, the incomes of citizens, the distances of planets from the earth, etc.

So *Fregean representations*, which are used in formal languages, are inherently incapable of *indicating their domains of interpretation*, because they always allow for unintended domains.

2.2.4. The argument of indistinguishability of reference

Now that it is clear that formal (Fregean type) representations cannot indicate their semantic domains: are they still capable of fixing the *exact reference*, that is the intended correspondence of thoughts and symbols *inside a given domain*?

The answer is *in general no*, and this is an important result of Hilary Putnam's theorem stated in *Reason, Truth and History* [134, pages 217–218]:

> "Let L be a language with predicates $F_1, F_2, ..., F_k$ (not necessarily monadic). Let I be an interpretation, in the sense of an assignment of an intension to every predicate. Then if I is nontrivial in the sense that at least one predicate has an extension which is neither empty nor universal in at least one possible world, there exists a second interpretation J which disagrees with I, but which makes the same sentences true in every possible world as I does."

Putnam basically says here that in model theoretic semantics, *truth of a sentence can always be maintained while its reference is changed*: "No view which only fixes the truth values of whole sentences can fix reference"[134]. If Putnam is correct, this means that there is always a second, different interpretation of a given theory in terms of a given domain, no matter how precise or detailed a theory is. No formal description will then be sufficient to determine reference within model theory. This is Putnam's main argument to reject model theory as a theory of meaning, and to propose that semantic reference is 'direct' — not 'indirectly' fixed via descriptions of properties, but *directly via acts of naming* (see also the discussion in Lakoff [90]).

The phenomenon is known in the ontology community under the name *'unintended models'*, mentioned already in Hayes' early account of ontologies in [65][11], and also in Guarino's foundational paper [56].

In the case of our *globe example* from Sect. 2.1, the reason why Putnam's devastating result does not affect the interpretations of the database query, is that those interpretations are coordinated by a spatial reference system.

2.3. Grounding data as a way of coordinating its semantic interpretation

We have seen that the solution to successful coordination is neither contained in natural science, nor natural language descriptions, nor formal theories. Looking at our example from Sect. 2.1, I suggest that the *competence of relating individual*

[11] "Indeed, no formal operations, no matter how complex, can ever ensure that tokens denote any particular kinds of entity"[65].

interpretations seems to lie in a *collective competence of humans* that has at least three parts:

1. the competence of referring to reproducible sensory-motor experience,
2. the competence of establishing a common 'observation language' about it
3. and the competence of expressing potentially ambiguous symbols in terms of this language.

The first and the second competence together enable the person with the globe and the person in Paris to refer to the same object, like the 'Eiffel Tower', and to refer to orientation concepts like 'north of', as well as to measurement standards. The third competence enables the person with the globe to 'ground' complex reference frames, like an ellipsoid or the globe, and thus to interpret calculations as operations on the experienced earth surface. For example, two points on a longitudinal circle on the globe can be interpreted in terms of the 'north of' relation.

We could say that the competence of humans to coordinate interpretations of a symbol is given by *operationalizing* that symbol, that is, to say what it means in terms of mental or physical operations[12].

It also involves the establishment of a primitive operational language about observation *as a social fact*. It needs the social re-construction of observation symbols by the act of pointing to something others can repeatably observe as well. It is just about what Quine called the 'edges of our web of beliefs which impinge on experience' or 'observation sentences', which are the basis for a method of grounding. Whether the web of knowledge as a whole is correspondent or not between humans is not in the focus and not required for this method. Also, it will not have to draw heavily on natural language with its multitude of usages and ambiguous notions, because it can introduce new symbols. And as the establishment of social facts needs clear and established methods of construction, symbol usage, and inference, a formal language is obligatory. But such formal theories do not determine interpretation, they are more an *inductive consequence* of a previously established way of using and interpreting the symbols.

[12]This view should not be reduced to a naive operationalism, which tries to restrict the meaning of every word to concrete instances of physical operations, which are often subject to change and non-repeatability. From the discussion above, it should also be clear that there is no computable strategy for deciding about the correspondence of operations and symbols, since it would again run into the symbol grounding problem.

Chapter 3

An operational view on perception, language reference and predication

Of one and the same thought, from mind to mind, by what means -
– through what channel – can 'conveyance' be made?
To no other man's is the mind of any man immediately present. [...]
'Under yon tree, in that hollow on the ground, lies an apple'.
At and during the time we are thus conversing, the ideas of the apple, the ground, and the hollow are in both our minds. In this way it is, that we learn the import of this same word 'in' with reference to our two minds.

— Jeremy Bentham [12]

Symbols are things that stand for other things. One may also say that they are a "sign" of other things. These other things are called *referents* and are usually of interest when trying to understand a symbol. How to refer to referents is called the *problem of reference*. The symbol itself is merely of conventional or formal interest. It may not even resemble its referent [34]. If symbols are formal in this sense, how can they convey what they stand for, in such a way that people know what others are talking about? It is clear that the word "convey" used here can only be a metaphor, similar to the image of a "conduit" [144]. Possible solutions to this problem will be in focus in this chapter.

As was argued in the last chapter, this problem reappears when describing the semantics of data. It lies, on the one hand, in the degrees of freedom in interpreting signs in terms of mental referents which stem from observation or imagination. On the other hand, it lies in the undeniable fact that there is no direct access to reality except via those referents. If information needs a *code* to be deciphered, as commonly assumed in information theory, then a cognitive organism needs direct access to this code, not only to the symbols of this code, but also to their referents and to the mapping between both. But a cognitive organism does not have access to referents beyond experience and imagination

[195, Chapter 6]. The main challenge for semantic engineering is therefore to adequately coordinate personal coding. Neither a fall back on natural science or on common natural language, nor an obligation to formal languages will accomplish this job (Chapter 2). The main methodical challenge for a semantic theory is to show how symbol interpretation can be made inter-subjective in spite of the 'informational closure' [195, Chapter 6], i.e., despite the fact that an inter-personal code needs to be realized in terms of personal experience and imagination.

In this chapter, I will suggest an *operational approach* to the problem of inter-subjectivity of language reference. The term language is used here in a broad sense, including not only natural languages, but also technical and formal ones. In the first section, I argue that such an approach could be based on perceptual operations that are constrained by the human sensori-motor apparatus (Section 3.1). In Section 3.2, it is argued that the human perceptual apparatus has the necessary invariance to ground the meaning of symbols. It includes Gestalt mechanisms that are autonomous, self-organizing and have bottom-up priority, while at the same time remaining top-down influenced by human attention, thought and will, so as to enable conscious experience. Section 3.3 discusses cognitive operations available for learning language reference. After reviewing current theories of language semantics, I argue for an operational view on language reference and meaning. Then I discuss principal operational sources, such as the apparatus of joint attention, and cognitive operations of predicating and referencing in natural language. I conclude in Section 3.4 with the proposal of a general view on the cognitive apparatus for language reference and predication.

3.1. Inter-subjectivity of symbol interpretation

Jeremy Bentham, a 19th century philosopher, had an influential idea about the kinds of things that language symbols denote:

> "Language is the sign of thought, an instrument for the communication of thought from one mind to another. [...] It may be the sign of other things and other objects in infinite variety, but of this object it is always a sign, and it is only through this that it becomes the sign of any other object." [12]

Bentham's view, cited from a book edited by C. K. Ogden, was constitutive for a famous approach to semantics called the *meaning triangle* (see Figure 3.1). An early one of these was published by Ogden and Richards [122]. Also, Frege's treatment of meaning [44] can be understood in this way, as well as many other approaches to the philosophy of language in the last two centuries[1]. The meaning triangle stresses the importance of an *intermediary thought* (the apex of the triangle) and *acts of reference* (the two upper edges) to connect symbols with their referents. The lower right corner, the referent, is usually taken to exist independently of human thought, but the triangle as a whole stresses that human

[1]Including Peirce, Carnap, Morris, de Saussure, Hjelmslev and Buyssens, see [34].

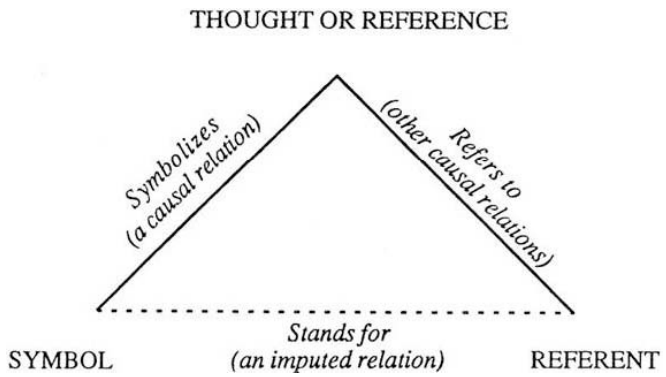


Figure 3.1: The meaning triangle with focus on its edges, cf. [84].

thought is needed to refer to it. The meaning triangle therefore stands in contrast to an objectivist approach to semantics.

The frequent use of the triangle may obscure that its use in explaining reference is limited. This is because a very important aspect of reference, namely inter-subjectivity, is realized across more than one of them. The meaning triangle leaves us in the dark about how inter-subjectivity of symbol interpretation is practically achievable, despite the apparent subjective degrees of freedom and arbitrariness involved (compare Chapter 2). Frege, for example, in his treatment of meaning [44], posits the existence of an *inter-subjective thought called 'sense'* (German 'Sinn'), distinguished from a 'referent' ('Bedeutung'). Frege's 'sense' is a shared thought that serves as a descriptive way (perhaps one of many ways) of referring to a referent. The most cited example is the pair "morning star" and "evening star", which stand for two senses to refer to one and the same object in reality, namely the Venus. In one or the other way, a meaning triangle approach seems to require *shared thoughts and objects* in order to account for the inter-subjectivity of language.

The story how we come to share thoughts remains mysteriously underexposed. On the one hand, how can a ('sense'-) thought be shared? We already saw in Chapter 2 that the existence of ideas in an interweaved network of knowledge is arbitrary to a large extent [136]. For a practical method of symbol grounding, it is not enough to posit such senses. Their assumption merely shifts the problem out of scope, as there may be many abstract thoughts that are not shared. On the other hand, it remains also unclear how a shared thought like 'evening star' can indicate a referent, like the Venus, in a reproducible fashion. Some philosophers like Searle have attributed humans with *intentionality*, i.e., the "magical" competence of referring to the world outside [164]. This allows them to maintain an objectivist view of semantics [165]. As Putnam [134] argued, and as we have seen in Chapter 2, there is no kind of representation that has the 'magical' intrinsic power of referring to things in the world. And unless one is willing to accept the problematic philosophical viewpoint of direct or naive realism, knowledge of

reality must be mediated via the private sensori-motor mechanism.

Sharing a referent could then perhaps mean to *share an experience*. But what does this mean if experiences are private, too? One could posit that sharing an experience actually means to *reproduce an experience by the private sensori-motor mechanism*. This is what I would like to propose in this chapter, and what has been suggested by constructivists like von Glasersfeld [195] and phenomenologists like Merleau-Ponty [113].

I suggest that reproducibility of experience can be physically realized in terms of constrained sensori-motor operations across a group of people in an environment. What Putnam's thought experiment of Twin Earth [133] shows, is, that extensional meaning is relative to an observer in an environment: The concept of 'Water' happens to be H_2O in our environment, but could mean XYZ on Twin Earth. But as von Foerster [192] has argued repeatedly, the biological as well as cultural invariances that are constitutive of experience are as well a product of this observer – environment system:

> "In contradistinction to the classical problem of scientific inquiry that postulates first a description-invariant "objective world" (as if there were such a thing) and then attempts to write its description, here we are challenged to develop a *description-invariant "subjective world"*, that is a world which includes the observer. This is the problem." ([192], emphasis mine)

Von Foerster's relevant insights are at least threefold: first, that higher level experiential invariances, e.g. the constancy of object and events, are invariances of representations that "emerge as consequences of branches of computation" [192]. Cognition as a whole then simply means 'computing a reality' based on some lower-level physiological constraints that are invariant across people. Candidates for such mechanisms are discussed in the following sections. Second, that these invariances, as well as the stability of organisms as a whole, ultimately arise as a stable dynamic equilibrium of action-perception cycles in a subjective environment[2]. And third, the apparent "objectivity" of observations - likewise - is the result of a stable operational equilibrium in a group of people [193].

If this is true, then a "description-invariant subjective world" needs to be constructed by a human oberserver based on operations of his or her sensori-motor apparatus. The operations involved are essential and cannot be easily explained or reduced to other "objective facts", since their stability is a product of complex biological and cultural evolution in an environment. For the purpose of a grounding theory, it is unnecessary to speculate about how these operations may be physically realized, and whether they are genetically or culturally determined. Instead, the fact that they account for the inter-subjectivity of reference could be enough to provide a *semantic plane* for information ontologies.

[2]See [193] and other authors on cybernetics. This idea can already be traced back to Jakob von Uexküll and his notions 'Funktionskreis' and 'Umwelt' [196].

Figure 3.2: Emergence in perception [94] (by kind permission of Steven Lehar ©).

This *operational view* on the inter-subjectivity problem does not force one to abandon the meaning triangle. As Kuhn has suggested [84], semantic engineering should be focused on the edges instead of the vertices of the triangle. These edges stand for the various ways how people observe reality, how they express thoughts through symbols and thereby refer to reality.

3.2. How conscious experience is based on Gestalt mechanisms

The most relevant cases of invariance in the human perceptual mechanism were recognized quite early by Gestalt psychologists [80]. These perceptual constraints give rise to some of the most puzzling mental capabilities of the human brain. They still seem to resist scientific treatment regarding both neurophysiological explanations and algorithmic modeling, as well as philosophical and psychological attempts to explain how perception works (compare the discussion in [94], Chapter 3). I will discuss these phenomena in the light of their function for conscious experience:

Perceptual grouping, emergence and reification. The most well known Gestalt phenomenon is Max Wertheimer's *perceptual grouping* by principles of similarity, proximity, good continuation, closure, symmetry and periodicity. Even though these principles seem simple, one should realize that it is very difficult to devise explicit rules that predict the grouping percept for arbitrary figures. The reason lies in the global configural qualities that seem to underly this mechanism [80].

The apparent holistic, or "global first" aspect of perception is most difficult to account for in terms of algorithmic imitations. It gives rise to the phenomenon that *figure-ground* distinctions *emerge* spontaneously by a certain configuration, as soon as the display contains enough detail, and as soon as the observer is

triggered to search for an instance of a certain Gestalt. In Figure 3.2, one may at first see only a structure of black and white patches. But as soon as one realizes that the image depicts a dalmatian dog snooping on the ground, the outline of the animal suddenly pops out of the image, as well as the ground with dappled sunshine under a tree. Inspection reveals that even non-existent edges in the picture (e.g. the back of the dog) can clearly be seen. This means that the visual system spontaneously fills in missing edges in correspondence with a Gestalt, which only partially "matches" the stimulus input. This phenomenon is called *perceptual reification* or 'filling in'. At the same time, a formally explicit description of the associated concept, e.g. in terms of explicit rules on patches of black and white, is hopelessly captured in a regress cycle: one would have to know which edges are spurious shadows and which are relevant for constructing the outline of the dog. But this distinction is exactly what cannot be made considering only local patches of black and white. So one would have to know the outline of the dog a-priori, which is impossible. If this a-priori distinction is impossible, then it may also be impossible to have a deterministic algorithm for computing the outline of the dog in arbitrary situations[3].

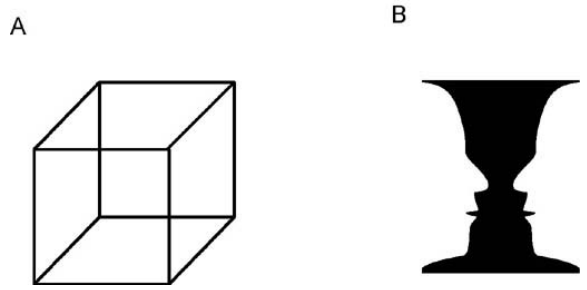


Figure 3.3: Multistability of the Necker cube and Rubin's face/vase illusion [94] (by kind permission of Steven Lehar ©).

Multistability. A famous example for multistability is the Necker cube, shown in Figure 3.3. Prolonged viewing of the image results in spontaneous spatial inversion of depth percepts that can hardly be controlled by conscious will. Multistability makes clear that there must be a spontaneous mechanism that jumps between discrete states of visual interpretation of an image. The important thing to note is that even though these interpretations are not stable, they are still *heavily constrained*, and therefore are an important source of invariance. For example, it is not possible two see both depth interpretations of the cube at the same time. This example shows that the conceptual and perceptual context together with a spontaneous Gestalt function realize a remarkable *bottom-up stability in*

[3]Note that it is nevertheless feasible to have an image segmentation technique trained at solving a particular example. But the general computational problem seems to be either intractable or has to be strongly constrained in variability [94].

spatial perception, which is quite independent of conscious will and discretionary influence of cognition.

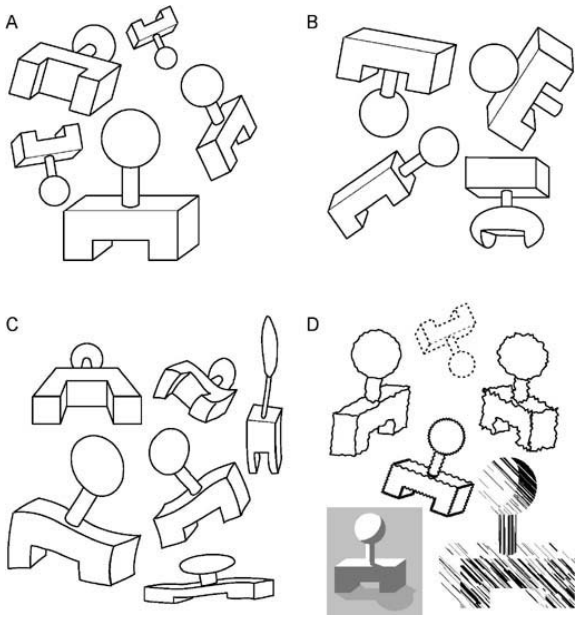


Figure 3.4: (A) Invariance of form under rotation and scale. (B) Different forms. (C) Invariance of form under perspective change and topological deformation. (D) Invariance of form under visual disturbance or brightness change [94] (by kind permission of Steven Lehar ©).

Invariance of form, color and relative position. An essential capability of visual perception is to recognize invariances of object properties: How the *form* of an object is easily recognized independently of scale, rotation, shift (Figure 3.4 A), perspective and topological transformation (Figure 3.4 C), as well as disturbances and changing brightness or illumination (Figure 3.4 D). Furthermore, different forms can easily by distinguished from same forms under transformation (Figure 3.4 B). In a very similar way, locations relative to surfaces can be easily recognized by a moving observer, for example the spot in front of a wall. Colors of objects, like the white of snow, can be identified independently of the color of the light falling on it at daytime or dawn. These capabilities seem to implement not only the recognition of form, color and position, but also the recognition of bodies through perspective change and through time. Similarly, relative positions are recognized as spots in the environment and easily tracked while moving around.

Amodal perception, spatio-temporal completion and imagination. This aspect of invariance is one of the most challenging ones for every theory of perception, since it implies perception in the absence of a stimulus. The fact that we perceive

occluded surfaces, e.g. the rear side of some body, as Gibson [50] suggested, or that we perceive the surface of the table on which the book is lying as continuing underneath [115], are examples for amodal completion. This mechanism allows humans to conceive of the world as a stable spatial configuration of distinguishable surfaces. Human attention often needs to be focused on this amodally completed layout, including the "world behind the head", as Lehar argues [94]. This can be demonstrated with every successful backward step [94]. In a similar way, the 3-D positions of visible edges and surfaces inside the perceptual field need to be reconstructed by the brain, as their relative locations are underconstrained by the 2-dimensional stimulus information reaching the retina. This is called the *inverse optics* problem (Figure 3.6). For example, the edges of the Neckar cube (Figure 3.3) are automatically perceived as standing in different angles and distances to each other depending on how the cube is spatially interpreted by a Gestalt mechanism. Depending on this Gestalt context, a visual edge may be reified as a corner or an apex or an occluding edge [94], and each of these interpretations spontaneously triggers the completion of the perceived locations of the adjacent surfaces. There are furthermore Gestalt completion mechanisms extending percepts into their temporal instead of their spatial dimension. For example, movements disappearing behind an occluder appear to continue, and two light bulbs flashing in close proximity are perceived as a continuous motion (*apparent motion phenomenon*). Finally, a very important but often disregarded human capability is the *imagination* of spatial scenes as occurring in dreams.

Every theory of perception has to explain these facts. Unfortunately, this has proved to be very difficult. The underlying mechanism has to be highly self-organizing and stable, while at the same time being sensitive to perceptual context as well as to top-down influences of cognition, as the emerging dog example suggests[4].

Regardless of any missing theory accounting for these facts, I argue that the mechanisms underlying these facts provide exactly the physiological backbone needed to allow for inter-subjective experience. This is because they provide necessary constraints across individuals for reducing the degrees of freedom in human experience. While addressing these mechanisms in the following sections, I will try to make as few assumptions as possible about the underlying functions. I will rather presume their existence and discuss the consciously available operations they give rise to.

Nevertheless, some *basic assumptions* about Gestalt mechanisms need to be made in order to make sense of Gestalt phenomena in the context of conscious experience and language. I will explain them now and argue for my view on the background of the contemporary discussion in cognitive science.

[4]We can see the dog as soon as we are searching for it.

We perceive an intermediary spatial reconstruction (ISR). Some form of preconceptual 3-dimensional spatial reconstruction of perceived space is necessary. In this representation, Gestalt completion is automatically computed, the relative 3-D locations of surfaces and edges are reconstructed, and Gestalts emerge *before* human attention is drawn to them and spatial reasoning, e.g. distance and direction comparisons, is performed. In contrast to the *direct realism* of Gibson [50] and other ecological psychologists, who claim that perception directly allows us to pick up features in the world, I assume that these features are Gestalts precomputed on some spatial reconstruction of the world viewed through an aperture. We never perceive the proximal stimulus, e.g. the image on the retina, itself. This is in line with Pylyshyn [135, page 111] and recent insights regarding human attention [161]. It follows the philosophical tradition of *indirect realism*. Arguments for this view will be discussed below.

Pylyshyn suggests that such a representation may be in active short-term memory. It must allow the observer to engage the motor-system by directly pointing to Gestalts and "reading off" 3-dimensional spatial comparisons like "above" or "behind". It also must allow to integrate spatial properties acquired through different modalities, e.g. auditory, tactile and proprioceptive sources of information [135, Section 5.4]. Furthermore, it must somehow account for the experienced continuity of space. I assume that this representation allows for spatial comparisons and the application of Gestalt mechanisms. But it should not be conceived as a spatial coordinate system with explicit point identifiers.

It is an open question how this representation is realized. Lehar [94] has suggested an analogue 'Gestalt bubble', an explicit spherical space realized in the visual cortex, into which images are projected and completed in 3-D via Gestalt mechanisms. The observer is located in its center. The perceivable depth is compressed nonlinearly towards the boundary, such that far things collapse with the 2-dimensional spherical surface, while near things become extended in depth (see Figures 3.5 and 3.6). Pylyshyn [135, Section 5] has argued that the internal representation of perceived space is not an analogue cortical space, but rather an internalization of external space. This space is equipped with objects identified by a Gestalt index based on natural constraints built into early vision (compare the FINST mechanism explained below). Therefore it is physiologically anchored in external space. The spatial reconstruction in this case is achieved in terms of complex visual-motor operations and transformations working on these indexes [135, Section 5]. This second approach may have greater difficulties in explaining how imagined space and Gestalt completion can be actually carried out, because here, internal representations are dependent on external stimuli.

Gestalt mechanisms are autonomous and self-organizing. I assume that if Gestalt mechanisms are applied to this intermediary spatial reconstruction, then they *spontaneously* generate or modify the configuration of Gestalts in it. This means that Gestalt mechanisms are not based on conceptual reasoning, even though

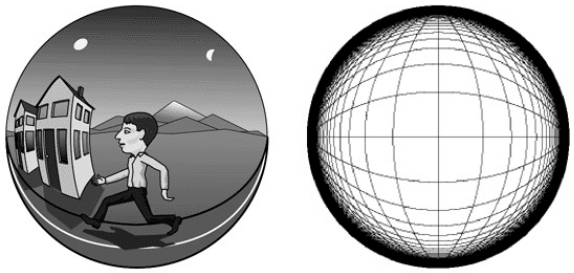


Figure 3.5: Steven Lehar's Gestalt bubble [94] (by kind permission of Steven Lehar ©).

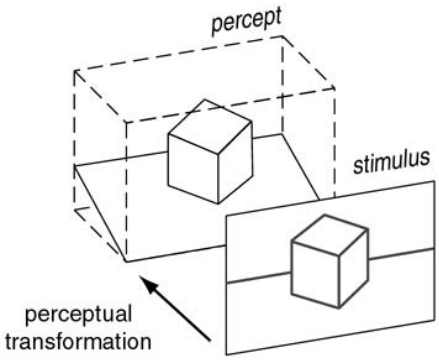


Figure 3.6: The inverse optics problem solved in the Gestalt bubble [94] (by kind permission of Steven Lehar ©).

their conscious application may remain sensitive to knowledge about a perceived scene. For example, if we know that the lower corner of the Necker cube in Figure 3.3 is in front, then the 3-D layout of a cube Gestalt is automatically snapped into the corresponding position without any conscious reasoning involved. Many of these mechanisms may even be triggered spontaneously - without any a-priori knowledge - by a certain configuration. How can we imagine a Gestalt mechanism? Köhler has argued convincingly [80], that in order to account for Gestalt facts, one needs to look for a spontaneous *self-organizing process* instead of a deterministic rule or concept. This process resembles an electric charge which distributes itself throughout a conductor, operating on the representation as a whole. It adapts partially to depicted features, but is also stable enough to overcome arbitrary missing facts, and fills them in if necessary. One possible candidate theory for such a mechanism could be Lehar's [94] *harmonic resonance theory*. Lehar posits an adaptive harmonic pattern of spatial standing waves realized in the visual cortex of the brain. Standing wave patterns, e.g. sound patterns or patterns of a vibrating steel plate, spontaneously adapt to partial constraints in their wave medium (for example singular clamped points on the steel plate). In this way, they can realize a variety of complex stable patterns, as can be seen in Figure 3.7.

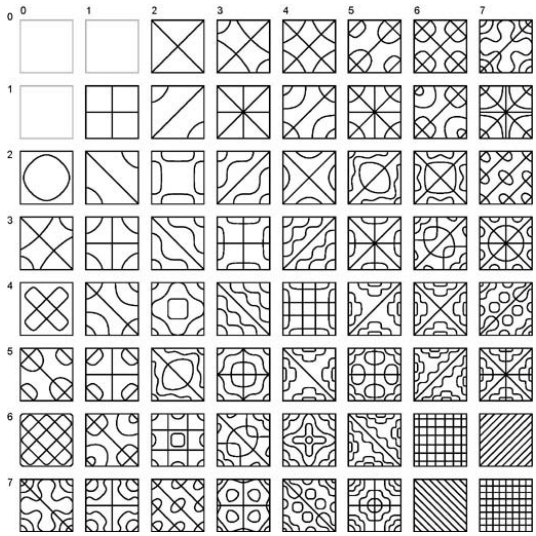


Figure 3.7: Chladni figures on a vibrating steel plate [94] (by kind permission of Steven Lehar ©).

Bottom-up priority of Gestalt perception over thought. A major bone of contention in perception research is the question in how far it is influenced by thought and in how far by the environment or physiology of the observer. Mainstream philosophy of the mind gives priority to thought (*'top-down priority'*). It says that *objects* (in general: particulars) are recognized on 'point-like' percepts using *conceptual reasoning* and the application of prior knowledge[5]. Likewise, cognitive scientists predominantly seem to think that experience of objects is a matter of projecting point-like stimuli, e.g. visual pixels, on the retina, and applying concepts based on known properties of those objects[6]. In this view, conscious experience beyond the sensory stimulus on the retina is mainly a question of knowledge. But such a view is problematic because it is at odds with the apparent stability, tractability and universality of some perceptual phenomena (compare the argumentation in Pylyshyn [135, Chapter 1 and 2]):

The tractability problem Pylyshyn [135, Chapters 1, 2 and 3] gives a list of convincing empirical arguments why perceptual reasoning processes would be intractable if individuals in a scene had to be identified by conceptual descriptions of properties. The general problem of determining object correspondences by properties – e.g. location and shape – in a visual scene is clearly intractable. Because of the 'binding problem' [135, Chapter 3.3.2], collections of properties alone lack the identifier of their host (properties are always *properties of something*). Therefore identification of objects is a

[5]See for example Strawson's influential idea of *feature-placing* [179].

[6]Compare traditional research in attentional studies [161], or the 'mental imagery' debate [135, Chapter 4.5].

necessary prerequisite for property detection, and not vice versa.

The missing leap from the world to knowledge It is then very difficult to explain how humans can share experienced facts. Since experience is based on knowledge, but knowledge is grounded nowhere, humans would be unable to refer to these experiences, including their experienced environment. Furthermore, it remains unclear how humans are able to share thoughts. It is therefore necessary that the conceptual network of knowledge is bottomed out in a purely 'causal' relation[7]. This contrasts with the relation of 'satisfaction' in model theoretic theories of meaning, such as Davidson's [30].

In contrast to the mainstream view, Marr [105] has pointed out the importance of subpersonal and unconscious *'natural constraints'* constitutive to knowledge and perceptual reasoning. These constraints must be hard wired into a part of the brain that is closed from conscious thought and therefore has *bottom-up priority*. This part was called *Early Vision* by Marr. Pylyshyn has argued for the existence of a pre-conceptual early vision process for *figure-ground indexing* [135] that should account for the causal connection to the world: FINSTs ('fingers of instantiation') provide spontaneous unconscious individuations of perceived objects. For example, the ball in the perceived scene of Figure 3.8 can be easily selected by attention just like an index in a file system. The top-down influence is still given by attentional selections of FINSTs and by guiding attention to scenes with certain concepts in mind, but not by contaminating object recognition itself by expectations and beliefs. The existence of such a mechanism accounts for

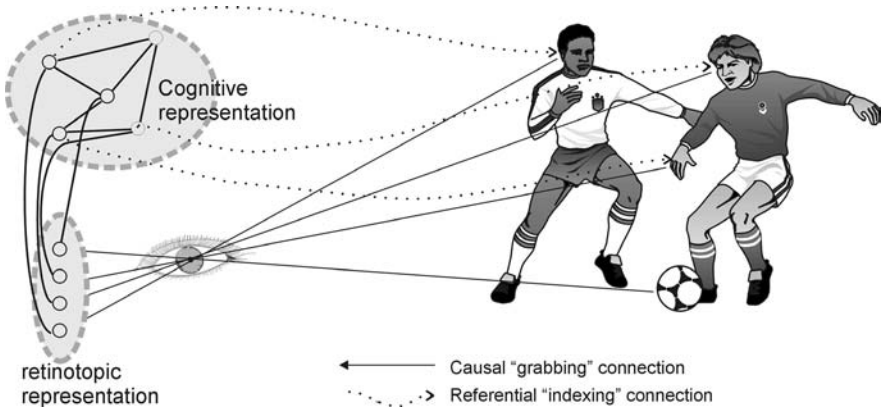


Figure 3.8: FINSTs enable to pick out and track several moving things even though there is not enough time to encode their properties [135] (by kind permission of Zenon Pylyshyn ©).

some of the discussed Gestalt phenomena. Interestingly, it is also at odds with

[7]With the term 'causal', I do not mean to imply a naive realistic 'caused be the world outside', but rather 'caused by lower level perceptual processes'.

the *physical symbol system hypothesis* [119], propagated by AI and mainstream cognitive science. It requires a bodily basis for cognition beyond mere stimulus input into a formal symbol manipulator.

Authors in the area of *cognitive semantics*, which at first glance stand in sharp contrast to Pylyshyn's tradition of thought, seem to have a similar understanding of bottom-up priority. In Barsalou's view, cognition and perception deeply interpenetrate, but perception prevails in case of conflict [5, Section 2.4.6].

Conscious experience means to focus attention on Gestalts. How does conscious experience fit into this bottom-up scheme? Consciousness can be explained as the product of attentional activity [161]. Barsalou argued that human attention can select patterns of perceptual states and store them in memory. These *perceptual symbols* can be used to reason with, or to simulate or reidentify things in a perceived scene [5]. However, the core content of perception resides at the level of unconscious neural representations, at which Gestalt mechanisms are located. Conscious experience in this sense can accompany, but is not a necessary ingredient of perception. Human attention is for example not necessary for many sensory-motor automatisms, while it is always involved in learning them. But once an aspect has been selected by attention, it has a high likelihood of being stored in long-term memory and being used as a fact of conscious experience in thought and reason [5]. Scholl [161] and Pylyshyn [135] have gathered many empirical arguments (I will present some of these later) that attention can focus on very complex Gestalts as well as on the space between them. Pylyshyn has argued that this is the causal link from sensory experience to thought.

In summary, I assume that thought and conscious experience is anchored in Gestalt perception in such a way that Gestalts spontaneously generated in an intermediary representation of perceived sensori-motor space build the anchor of reference for human attention. What *kinds of Gestalts* humans can attend to determines their range of perceptual capabilities, and thus the range of shareable experiences. I will make some suggestions about this range in Section 5.1.

3.3. How language and thought adheres to perception

Language and thought are usually conceived as the two essential *faculties of the mind* that give rise to human intelligence. If language epitomizes thoughts but expresses them only to a certain degree, then what else does it take to understand these thoughts? Is language the ultima ratio of thinking, as Wittgenstein and many philosophers after the linguistic turn held [207], or can we get around it? Lorenz and Kamlah [97] have argued that while we cannot get around the human cognitive capacity underlying language, we can in fact always dispense with a particular language, whether it may be technical or natural. It is important to realize that what it takes to learn a new language is not possessing a language, as

a child can do so without. Language therefore requires something beyond itself in order to be learned and understood from scratch, and thus a particular language is in fact dispensable. As I argued in Chapter 2, language as a communicative tool requires operations for inter-subjective reference that need to go beyond mere syntactical rules. In this section, I will keep track of such operations in contemporary work in language philosophy, cognition and linguistics.

The *problem of language reference*, i.e. what kinds of things sentences or words stand for, seldomly shows up in situations of natural language speech. This can be explained if we assume that involved operations were already internalized by the speakers, so they are not aware of them. Therefore, it might be insightful to analyze the problem from the perspective of a *language learner*. If one is confronted with unknown datasets, the situation is very much like learning a new language without a bilingual dictionary: The data consumer resembles the language learner in being often unsure about the adequate way of interpreting data expressions. I will treat the issue therefore from the viewpoint of learning an unknown language without a dictionary. A similar approach is also taken in Quine's *Word and Object* [143].

I will first discuss how current theories about language semantics cope with the problem of reference. I will then introduce an operational view on the problem which differentiates two kinds of operations, namely *predication* and *referencing*[8]. I will argue that learning these requires a common experiential basis (Section 3.3.2). Lastly, I will discuss the general cognitive apparatus required to learn predications and referencing in natural language (Section 3.3.3).

3.3.1. Language reference and contemporary semantics

The dominant ways of thinking about language semantics can be summarized under the headings of *objectivism*, *nativism*, *conceptualism* and *cognitivism*.

One classical approach is *objectivism*, which ranges from Frege's objective notion of 'sense' to the idea of *truth-conditional semantics* underlying Davidsonian [30] or Tarskian approaches. In objectivist approaches, linguistic expressions denote objective mind-independent entities [93] by truth conditions. But objectiveness is a metaphysical credo that can be believed or not. And what is more important, truth conditions of logical sentences cannot fix reference, as we saw in Sections 2.2.3 and 2.2.4.

Nativist approaches, like the one of Fodor [40], place concepts into the mind of human beings, but share the objectivist idea that these are mind-independent (intentional realism) [68] because they are *innate* in the sense that every human being is genetically equipped with them. In consequence, nativist approaches tend to neglect *productivity and creativity*, the human ability to create new concepts that have never been thought of before [68].

[8] "Referencing" in linguistics and philosophy is a more special term than "symbol reference" dealt with so far, as it is normally restricted to individuals denoted by nouns.

Conceptualist approaches, like the one of Jackendoff, posit therefore a comparably small set of innate concept primitives from which an infinite variety of concepts can be generated via grammatical formation rules. Lexical concepts for words like "John", in contrast, need to be learned in terms of so called 'semantic fields'. Among these fields are the continuous domains of "spatial structures" of objects, actions, or focal values such as colors, but also more abstract domains like "possession" [68]. Each such field comes with its own innate primitives, construction and inference rules. But most of the lexical concepts are opaque to any analysis in terms of syntactic structures, which is why they seem indefinable [68]. Semantic fields closely parallel the idea of *conceptual spaces* put forward by Gärdenfors [48]. In the following, the general idea of such semantic spaces or fields is reflected in the term *mental domain*.

The linguistic phenomenon of polysems appearing in metaphors, as well as in metonyms and prototypical categories, inspired *cognitive linguists* like Lakoff [90] and Johnson [74] to assume an *embodied and analog Gestalt* basis for all concepts. The idea of Johnson is that the concept which reflects an assertion is not a *proposition*, i.e. a mental statement about the world, but a Gestalt which bears a propositional aspect [74]. These Gestalts are called 'image schemas'. While they stem from the spatial domain of experience, they can be applied metaphorically to abstract domains. They can also be combined to form new Gestalts and to allow for inferences. Lakoff [90, Chapter 17] has claimed that image schemas constitute a *foundational level of thought* which is "directly meaningful".

What do these latter cognitivist and conceptualist approaches offer regarding the problem of reference? On closer inspection, they have severe difficulties in explaining how language reference is learned.

First, because image schemas are employed both in order to understand *and* link different domains, they hover in a rather *abstract realm* which is a compromise between concrete domains genuinely understood. Thus, they cannot account for the identity of structures in these concrete domains. Consider such diverse mental domains as movement and possession metaphorically linked by a *path* image schema like in "the book goes to Henry". These two domains exhibit quite specific structures on their own which are not captured in the path schema. For example, sensori-motor space is 3-dimensional including concepts like upward and downward, but possession is not; sensori-motor space is nearly continuous, whereas possession is discrete, which is the reason why one cannot say "give the book toward Henry" [69].

Second, while spatial metaphors are useful, it seems problematic to *reduce mental domains to the spatial domain* as Lakoff does [90]. Jackendoff argues convincingly [69], that even though the spatial experiential domain has a central role, concepts cannot be reduced to it. For example, the possessional concept "give" in "Bill gave a book to Henry" cannot be reduced to something like "Bill took the book over to Henry", because possession involves more than a spatial movement [69]. Image schemas, which are mostly inspired by the spatial domain,

do not seem to be sufficient to understand a new domain because they are devoid of indicators for referencing structures in this new domain. Jackendoff is therefore right in claiming that there must be special primitives available for constructing abstractions like ownership and right [69].

Third, Jackendoff's own claim that these primitives must be also abstract and non-experiential [69] seems to be problematic in its own right. The reason has to do with the problem of how to refer to these presumed innate abstract concepts, for example the concept of possession, if they are not somehow given to the senses in commonly perceived situations. For the same reason, Gärdenfors' conceptual spaces [48] as abstract mental domains may require grounding in perception in order to be learnable.

3.3.2. An operational view on meaning and reference

In deviating from the approaches cited above, I suggest that learning languages means to *refer to operations based on perception*. These operations have *predicative and nominative parts* for generating facts from Gestalt perception and *constructive parts* for generating ideas based on these facts. Language learning requires that a learner and teacher have access to such operations, as well as practical ways of disambiguating between them in jointly perceived scenes. This is regardless of whether these operations are products of genetics or culture.

The operational view proposed in the following has several intellectual ancestors. It adopts von Glasersfeld's view on meaning as an act of interpretation in terms of constructed mental entities [52]. It recognizes speech acts as a type of inter-subjectively available communicative operation [163], but goes beyond in identifying *predication* and *referencing* as the central kinds of acts that need to be learned in order to account for language reference. In this, it resembles operational theories of language meaning such as the one of Wilhelm Kamlah and Paul Lorenzen [76], or the one of Kuno Lorenz [96][9]. Like Lorenzen, it also recognizes *term construction and abstraction based on predications* as further kinds of operational competence [100]. It deviates in that it focuses on operations for referencing and predication as distinguished from meaning, and considers different means of abstraction (see Chapter 4). Furthermore, it takes *perceptual operations*, based on human attention and Gestalt perception, as indispensable requirements for learning predication and referencing. In this, I follow an argument originally given by Quine in [137].

Meaning is a speaker's act of interpretation. *Meaning* is not considered to be a concept or thought, nor a mind independent thing. The term meaning can be used as a verb in any sentence of the form "The speaker means ...", which denotes a certain *act of the person who utters a sign* [84]. If we say that the

[9]Lorenzen, Kamlah [76] and Lorenz [96] use the terms 'Prädikation' for predication and 'Nomination' for referencing.

speaker means something with this sign, then we want to say that we suppose him or her to *interpret* this sign in terms of his or her own thoughts or concepts. Furthermore, if we say a listener understands a sentence uttered by a person (a speech act), then we suppose him or her to impute meaning to the speaker's words, i.e. we suppose he or she presumes that the speaker interprets the sentence involved. The listener *understands* what the speaker says if he or she performs a corresponding act of interpretation on his or her part.

Interpretation (construction and association) cannot be perceived. I follow von Glasersfeld [51, 52] in assuming that interpretation involves an act of association of an *experienced sign* with a *constructed mental entity.* Interpretation requires that the linked mental entity *can be constructed*, because it may not be present in mind. In these cases, construction may need to be learned first. So there are two degrees of freedom of interpretation: one is the creativity to construct mental entities, the other one the freedom to associate a sign with whatever mental entity is present in mind. The problem of semantic heterogeneity is basically that this act of interpretation is *private and unobservable* for others. Therefore, *meanings in the sense above cannot directly be shared* among speakers of a language. And this is why speech acts and perceptual operations are needed.

Speech acts can be perceived. A *speech act* is an intentional human act of communication, first introduced by J.L. Austin, and further developed by pragmatic language philosophers like Grice [55]. John Searle, and likewise some constructivists [96, 76], considered it as the building block of linguistic communication. It can be shared in a certain sense. A speech act is a complex operation that was characterized by Searle [163] as consisting of 3 parts,

(3.1) $F(P(R))$ "I claim that ((Peter) is 24 years old)"

where F denotes the *illocutionary act* (expressed by "I claim"), R stands for the *act of referencing* (expressed by "Peter"), and P for the *act of predication* (expressed by applying the predicate "is 24 years old" to Peter). The act of referencing is sometimes also called naming. Searle suggested that every intentional production of a sign involves some of those acts (at least an illocution), but not each act has to be expressed via a perceivable symbol or sentence. For example, an illocution, like a question, command or claim, may be inferred from the communication context, so that sometimes a gesture on the part of the speaker expresses a whole speech act.

One may be tempted to conclude that this is the end of the story, that meanings are perceivable as part of the perception of the illocutionary and propositional parts of a speech act uttered by means of certain symbols. But this would oversimplify what is going on. If I perceive a speaker purposefully referring to something and predicating something with the help of a symbol, it does not follow that I can perceive what entities he or she is referring to or what kinds of

facts he or she is hinting at. The listener needs to *imitate the underlying interpretations*, even though large parts of them may be hidden from the perspective of the listener. The subtleties and challenges of learning predications were first discussed by Quine in [143], and are analyzed by Lorenz in greater detail [96]. I will analyze them more deeply in the remainder.

To see the contribution of speech acts clearly, I suggest that they give hints as to what *kinds of interpretations* the listener is supposed to do in order to imitate the speaker, i.e., to understand the utterance[10].

Learning referencing and predication involves perceptual operations. The language learner's freedom to associate an expression with a concrete mental domain must be effectively restricted by the teacher. In such a domain, object names, deictic words, as well as so called definite descriptions [147] must be linked to identifiable entities, and predicates with concrete operations on them. My argument is that, unless in case of mathematical domains, the learner's choice can only be effectively restricted by operative means of perceptual experience[11]. This is regardless of whether learning involves what Russell [147] called *direct acquaintance*, or whether experience enters the stage at a more subtle level.

Searle suggests that the *act of referencing* involves that the speaker must be able to identify one and only one object for the listener, so that the listener can easily "pick out" what is being talked about (axiom of identification) [163]. This is done by one of the following *definite referring expressions*:

1. *proper names*, e.g. "Sokrates" or "Russia" (as distinguished from names for universals or categories)
2. *definite descriptions*, e.g. "the highest mountain of the world".
3. *pronouns and deixis*, e.g. "this", "that", "I", "it" .

In order to make use of these expressions, the listener of course already has to understand them. Understanding *deixis* depends on the perceived context in which they are uttered. *Proper names* can be used to refer to objects when they are not currently present to speaker and listener, but need to be learned via direct or indirect perceptual acquaintance, as Russell suggested [147]. According to Quine, they can be learned based on a combination of a pre-conceptual similarity basis[12] and ostension [137].

[10]Searle seems to underestimate this problem in presuming that his notions of 'shared intentionality' and 'semantic background' could account for it [163].

[11]Note that the necessary restriction may also be impossible, because the intended domain supplies no link to experience. In this case, learning referencing or predication is presumably impossible. But note that the learner is of course still able to understand a proposition. My point is that this understanding cannot be effectively coordinated on the part of the teacher, as he or she lacks vital resources for teaching the restrictions involved. The result is uncertainty in interpretation.

[12]Quine does not say very concretely what he means with this term, but it is clear that he thinks about some pre-conceptually available sensory mechanism. This may be e.g. a Gestalt.

If the proper name is not learned via perception this way, then it must be learned using a predicate, e.g. a definite description, that relates this name to other known entities. *Definite descriptions* are of the form "the brother of John". In this example, the noun phrase expresses a predicate which already contains another referring expression, e.g. "John". The learning of the definite description therefore is based on learning the predicate and this name first.

How about predication? Searle is not very explicit on this, but Quine has more to say here. He suggests that, similar to learning a name like "Fido" by pointing at a dog and learning a (Gestalt) criterion of perceived identity for it, we learn the general term "is a dog" by pointing at different dogs and learning a *perceptual similarity* across individual dogs [137]. Similarly, Lorenzen [100] and Kamlah [76] suggest that predications are learned by what they call *exemplary introduction* ('exemplarische Einführung'), i.e., by demonstrating examples and counter-examples (see Figure 3.9).

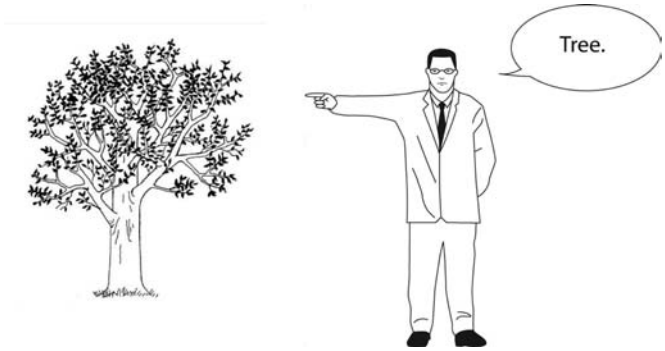


Figure 3.9: A predication performed by ostension.

All possibilities so far involve perceptual operations. But are there also other ways of learning available?

One may argue that there is a way of learning predication and referencing not via perception, but via *axiomatic theories*. However, there is a convincing argument against this view, which is a variant of the indeterminacy results in Sections 2.2.3 and 2.2.4. A related idea stands behind Quine's thesis of *indeterminacy of translation* [143]. In Appendix 8.1, I have illustrated this argument by a modified 'Chinese Room' scenario, compare [164].

Nevertheless, there is in fact one way of learning referencing and predication without perception. It is the way how mathematical abstractions like transfinite cardinal numbers [99] or n-dimensional geometry need to be learned. But this possibility seems to exist only and uniquely for *mathematical domains*. As Piaget [126] and Lorenzen [100] have argued, and as I will explain in Section 4.1.2, numbers and arithmetics can be learned as abstractions from constructive mental operations themselves, e.g. as class abstractions from recursively specified countings [100]. The important observation is that these abstractions depend only on

certain mental operations which are non-perceptual. For example, number classes can be identified without reference to concrete counted objects, as long as the operation as such is at one's disposal.

The more general insight behind these considerations is that learning referencing and predication needs to be based on other mental operations, not on axiomatic theories, and that in the case of empirical domains, perceptual operations are necessarily involved.

Gestalt references can be disambiguated by guided attention. The previous claim stands or falls by the subtle question why and in how far perceptual domains are better suited for learning referencing and predication than formal axiomatic theories.

One could argue against this, claiming that perception and ostension are themselves indeterminate. Actually, the very same argument above was used by Quine to show that reference and predication learning are *indeterminate* even in case commonly perceived scenes and ostension are available. He illustrated this in his famous example of how to learn the meaning of 'Gavagai' by pointing at a rabbit [143]. The mere pointing allows teacher and learner to link their attentional focus in a scene. But it does not rule out that 'Gavagai' could either mean the general term 'rabbit', the one individual rabbit, a spatial part (its leg), or a temporal stage (the rabbit now), or an action (the rabbit hopping). The problem is that *there is more than one potential Gestalt in focus*[13].

This puzzle has had major influence on the philosophy of the 20th century after it was stipulated by Wittgenstein in his *Philosophical Investigations* [207]. It was involved in the abandonment of empiricism as a philosophical approach to knowledge, because it implies that there is no unambiguous way of translating sentences into direct experience [136]. But it is important to note that it was Quine himself who suggested that children successfully learn referencing through that very same empirical mechanism [137, 139, 140]. What seems to be easily overlooked if only reading 'Word and Object' [143], is that Quine actually pointed to the practical solution of this problem in terms of what he called *observation sentences*. An observation sentence is learned using ostension in a language game of shared attention. This game is based on perceivable similarities [137] without appeal to a-priori beliefs on the part of the observer [140]. The point I would like to make is that the use of such language games on the basis of a range of perceptual similarity criteria (given by Gestalt mechanisms) usually allows a teacher to rule out unwanted interpretations, and thus, as Lorenz [96, page 209] claims, to successfully teach the underlying operational schemes.

For example, because humans share a Gestalt mechanism for individuating

[13] In order to make this challenge more apparent, suppose that by pointing at a statue someone utters "Posąg" (polish for statue). Does this person mean the amount of clay, or the human artifact? And by pointing at someone drinking coffee: does this person mean the movement, or the intentional act of drinking coffee?

bodies, they can distinguish bodies from each other (see Section5.1.3). By pointing at one, the teacher can rule out the possibility that the learner associates the wrong body. Since only a certain kind of a body, namely a rabbit, is in focus in the Gavagai scene, the pointing also rules out that 'Gavagai' means elephant or another category. Furthermore, whether 'Gavagai' means the general category rabbit or an individual rabbit can be easily determined by pointing at another rabbit and asking 'Gavagai?'. If the answer is yes, then 'Gavagai' must be a general term. If the term can be applied perpetually but only to a single body, it is a name. Also, the possibility of interpreting 'Gavagai' as rabbit parts or stages can simply be ruled out by pointing at other parts at other opportunities asking the same question. This argument will become clearer when discussing Tomasello's work in the next subsection.

3.3.3. Cognitive apparatus for learning referencing and predication

If the assumptions of the last subsection are correct, there must be *operational sources* to disambiguate and learn referencing and predication. These should be available in every mental domain humans can refer to with the help of language. I will develop here a broad view on the underlying cognitive apparatus in central domains of human cognition. Drawing on empirical results in cognition and linguistics, this subsection illustrates that tackling with the problem requires a comparably simple apparatus of human attentional control, perceptual Gestalt mechanisms and mental construction.

3.3.3.1. Operations for joining attention and perceiving intentions

Based on a range of empirical support, Michael Tomasello [188] proposed that the major distinctive feature of human cognition in comparison to apes, showing up in children of 9-12 months, is their capability to conceive of other bodies as *intentional actors*, i.e. as purposeful and goal-driven beings, similar to themselves. This means that they begin to interact with their environment in a *triadic* instead of a dyadic fashion (see Figure 3.10), enabling the phenomenon of *joint attention*. Joint attention is unique to humans, because it requires (1) to perceive all involved actors' intentions as directed to something, (2) to mutually follow or draw the other's attention to something, (3) to perceive other's *intentions to draw one's own attention* to something [188] and (4) to exchange roles in these intentional actions.

According to Tomasello, this unique capacity has a range of important consequences for human cognition:

For example, it enables what he calls *cultural learning and the cultural evolution of language*. Humans can *imitate* the intentional behaviour of others, and so they learn to regard things as *cultural artifacts* with an intended function shared with others. The whole story of human cognition can be told in terms of cultural learning (history) and piecewise mental innovation (ontogenesis), triggered by

the innate ability of joining attention, imitating intentions and constructing new mental categories. In particular, sharing a language, as Mead had argued earlier [112], can be seen as a special case of imitation, involving not only reproduction of utterances, but also reproduction of intentional attitudes. It is comparable to the evolution of chess, having a large inherent indeterminacy and freedom, constrained only by a small set of cognitive abilities that allow to learn and modify the game and to pass it on [188, Chapter 7].

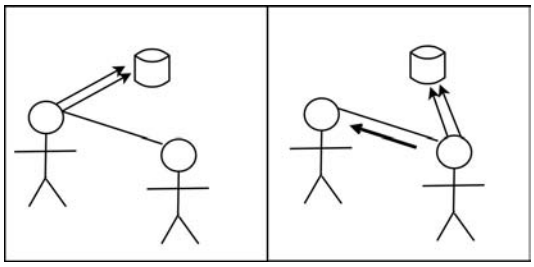


Figure 3.10: Two types of intentional interaction: following someone's attention to something (left, thin arrows) and drawing someone's attention (right, thick arrow) to something, cf. [188].

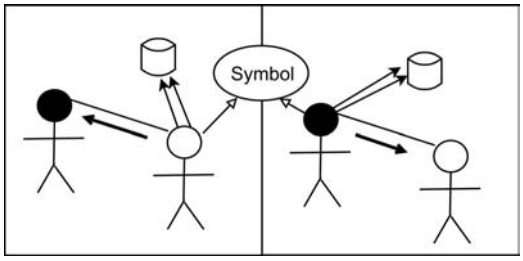


Figure 3.11: A communicative act (symbol uttering) is a special case of mutually drawing the attention to something, cf. [188].

Communicative acts like speech acts are just a special case of joining attention, namely an act of drawing someone's attention to something using conventional symbols (Figure 3.11). The same mechanism explains how a learner is able to imitate *acts of referencing and predication*. These are tricky cases of joint attention, but they are mastered by 2 year old children in appropriate language games. Tomasello describes such games in Chapter 4 of [188], and thereby illustrates that proper names, as well as "verb islands" like "give", "bring" and "kick", can be learned via joint attention, even if word utterance and referent never co-occur in a situation. For example, teachers were able to bring children to interpret fictive words as either action verbs or object names, depending on whether either a certain action or an object was new or absent in a repeated scene whenever the word was uttered[14].

Via the mechanism of joint attention, inter-subjectivity of language names and predicates spreads towards other fields of cognition. It thereby creates a plethora of *human perspectives* on the same scene which is typical for the semantics of human languages [188]. For example, humans can not only perceive what kinds of actions the environment offers to them, *but also what it may offer to other persons* [188, Chapter 4].

[14]For example, an adult uttered "And now Modi" while throwing objects down a pipe. If the adult afterwards introduced a new kind of object, children would interpret "Modi" as denoting this object type, but if a new action was introduced, they would interpret it as a name for the new action type instead.

In summary, I argue based on Tomasello's work, that humans have the distinctive capability to realize *joint attention*, which enables them to teach and learn predication and referencing. In order to do so, humans need to *perceive other's intentions*, and this requires them to understand and anticipate all kinds of *human actions* for themselves as well as for others. Learners of natural languages must be equipped with appropriate capacities.

3.3.3.2. Operations for predicating and referencing in natural language

Humans are endued with a cognitive apparatus that creates the *bunch of perspectives* from which joint attention picks out certain aspects. In this subsection, I will discuss the role attentional perspective plays in natural language semantics, based on suggestions made by Leonard Talmy and Ronald W. Langacker. This allows to develop a comparably simple view on a mental operational device whose existence is supported by a wide range of linguistic evidence.

The cognitive linguists Langacker [91] and Talmy [183] argued that the grammatical part of language, as much as the lexical part, involves meaning in terms of embodied cognitive operations. In their view, verbs and nouns have an operational semantics over and above the concrete lexical content. Unfortunately, their suggestions are far from being concise, and thus need to be clarified themselves.

The unifying simple view I propose is that the referencing and predicating parts of language sentences, roughly identified with its noun (NP) and verb phrases (VP), are really instructions on moving attention to some entities in a mental domain (NP), relating them via Gestalt mechanisms (VP), and constructing new domains from such relations (NP). This view will be explicated in the following and lined with ideas from Langacker and Talmy. The central role that attention plays here has been suggested before by some constructivist philosophers, like von Glasersfeld [195, 194], and was recently applied by Marchetti [103] to suggest a similar unifying view on Talmy's cognitive semantics [104].

Control of perspective and attention. If the propositional part of a sentence always implies a fact, as was traditionally held by philosophers (compare Russell's view [147]), then what kind of fact may this be? Traditionally, the answer was a state of affairs in the world. But this stands in apparent contradiction to some *fictive*, *dynamicist* and *perspectival* linguistic phenomena, that suggest that those facts are rather *construed by observation* of an actual or imagined (fictive) situation.

Take for example the phenomenon of *fictive motion*: [184]:

The landscape spreads towards the horizon
*The spilled soup spreads on the floor
The cliff faces towards the valley
*He walks towards the valley
The scar extends from his ankle to his knee
The scar extends from his knee to his ankle

Here, the non-fictive cases are marked with an asterisk. In the fictive cases, a speaker does not believe that there actually is a moving entity involved, even though understanding the sentences requires the assumption of some implicit movement. According to Talmy [184] and Langacker [92], this movement is a consequence of our general way of *constructing stationary circumstances* via fictive motion in some mental space. The semantics of such sentences is given in terms of *mental scanning* of an imagined or observed object, and in terms of summarizing experienced properties along the scanned path in order to construe a complex configuration (called *summary scanning* by Langacker [92]). The direction of movement thereby accounts for the semantic difference of the last couple of sentences, which are not distinguishable by objective fact. Similarly, human attention accounts for the phenomenon of *perspective* [183] and is used in order to *distribute attention* [185] across an actual or imagined scene.

As Langacker [92] suggested, all these kinds of linguistic dynamicity stem from fundamental cognitive operations on mental domains which naturally extend to *non-spatial, abstract or virtual situations*. For example, the scanning operation can also be used to *move through time* in sentence a), or to scan through a fictive domain of options for actions in sentence b):

a Through the centuries, we have had many great leaders
b Each option seems worse than the last

Marchetti proposed [103] that this kind of dynamic cognition can be conceived as an *attentional operation supported by unconscious counterparts*. The unconscious counterparts thereby help to structure and construct perceptive (tactile, visual, proprioceptive), interoceptive (pain, pleasure), intentional (goals, actions) and abstract mental domains, on which human attention can then be focused in order to produce the plethora of conscious experiences [195].

Selecting and relating entities with attention. Language speakers make use of the apparatus for controlling attention mainly in order to select entities, and in order to disclose and record structural properties of a domain in focus.

Langacker [91] assumes there are operations on a domain ('base'), which focus attention on one of its substructures ('profile'), while a larger part of that very same domain is brought into broader view ('scope') in order to invoke the substructure. For example, the noun "island" *profiles* a land mass which needs to be bounded in a *larger scope* of the surrounding sea. Scoping and profiling are realized via attentional and perspectival control discussed above, but in addition there is an operation to disclose and record the *structural properties*, e.g. the boundedness of the island.

Langacker distinguished *relational* from *nominal predication*, the former of which profiles *interconnections* in a domain and roughly corresponds to predications, while the latter one accounts for complex unities referred to by noun phrases and corresponds to referencing. For example, the noun "group" profiles

an entity like in Figure 3.12, which can be analyzed in terms of relational predications like "together". Nominal predications often imply the construction of unities in terms of relational ones, i.e. to construct "groups" from a "together" predicate. Relational predications are realized via *summary scanning* [91]. Summary scanning means that a configurational property is built up cumulatively in memory while scanning part of a domain.

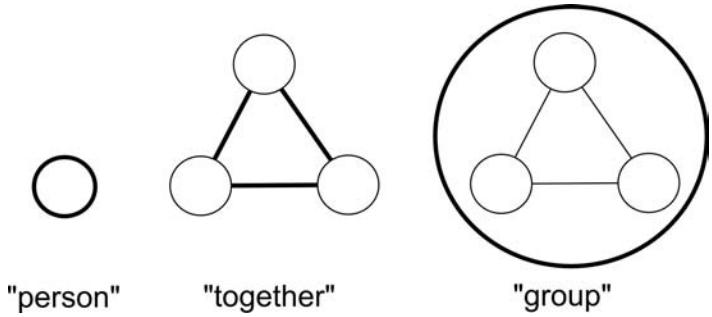


Figure 3.12: The interplay of interconnections and nominal predication for the noun "group", cf. [91].

In order to account for the *structure* which is recorded by mental scanning, Langacker [91] proposed only a single type of mechanism called *boundedness*. Similarly, Talmy [183] argued that the structure expressed by grammatical notions in general is *topological* in nature, rather than Euclidean. Interestingly, contemporary research in spatial cognition provides evidence for viewing event as well as object cognition to be based on topological structures [79]. This kind of structure together with attentional scoping and scanning seems to suffice for explaining many grammatical phenomena in terms of cognitive operations.

For example, the idea of a bounded region allows Langacker [91] to capture the semantics of *count vs. mass nouns*: count nouns directly profile a bounded region in some domain, while mass nouns profile an unbounded region in the current scope of predication:

a I see a red spot
b I see (nothing but) red
c The furniture consists of single chairs
d Each house comes with a bulk of furniture

In the first sentence, an observer's scope of attention includes the surrounding white wall, on which the spot is painted, while in the second sentence, the observer is just too near to see the boundary. In a), "spot" is a count noun, whereas the mass noun "red" in b) is construed by zooming in and loosing the boundary out of sight. The meaning of mass nouns can also be construed by zooming out and profiling unbounded pluralities of bounded entities, as in the case of "furniture" in sentence c). Furthermore, new kinds of count nouns can be derived from mass

nouns by introducing new kinds of boundaries at another scope level. This was done for the count noun "bulk of furniture" in sentence d).

In summary, it seems that linguistic domains of predication are mainly structured by a simple kind of Gestalt, namely a kind of bounded region. Language serves to build attentional perspectives on such regions in order to pick out and interrelate entities via scanning. These relations can then be used to construct more abstract entities (referred to by nouns, see next paragraph). I suggest that Gestalts *prestructure* domains and therefore *allow to establish interconnections*. Depending on the scope and focus of attention, boundaries come into sight or not. Therefore interconnections are either constrained by boundaries or by the current scope. The advantage of this view is that it is simple and that it allows attentional as well as pre-conceptual and unconscious mechanisms, namely Gestalts, to enter the stage of language cognition.

Constructing domains of predication. The competence which is in focus here, is the competence of *constructing domains* that were not present before (i.e., conceptual productivity). Langacker and Talmy suggested language tools to induce new domains of predication. These include:

1. *Unification*, i.e. triggering the construction of *bounded or unbounded unities* of interrelated entities. This operation is involved in nominalizations, as e.g. "jump", "explosion" (bounded) or "breathing" (unbounded). Bounded and unbounded unities are constructed based on *summary scanning* and attentional scope from a structured domain [91], as discussed above.
2. *Boundary introduction* in a given domain, in order to yield new unities. This was called *portion excerpting* by Talmy [183], as in "stand of timber" or "breathe for 3 hours".
3. *Multiplexing*, i.e. discrete repeating of a bounded region in a domain in order to yield a new unbounded unit, as in "birds flew in" or "keep jumping" [183].

Combining these operations gives rise to different kinds of unities. These can then be referenced by new count or mass nouns. The kinds of unities correspond to configurations proposed by Talmy [183] (see Figure 3.13). The unified entities are held to populate new and more abstract domains, called 'planes' by Langacker [92].

The ideas outlined above provide a rather comprehensive view on cognitive means involved in referencing and predication. But they remain to be spelled out in many respects. For our purpose, it is necessary to focus and concretize them. In this thesis, I will conceive predication and referencing in terms of attentional selection, Gestalt application (boundary introduction), scanning of Gestalts and constructive unification. Unification as constructive means turns out to be particularly useful, and gets a formal treatment in Chapter 4.

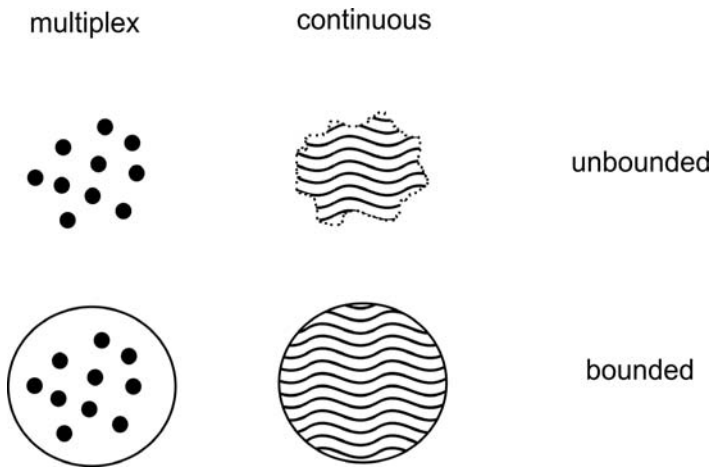


Figure 3.13: Types of unities for constructing new domains based on Talmy's configurations, cf. [183].

3.4. Cognitive operations for referencing and predication

In this section, I will sum up the results of this chapter and propose a general view on the cognitive operations available for referencing and predication. Regardless of how natural language cognition is physiologically realized, the types of operations involved are at our disposal for learning all kinds of languages. As Lorenzen argued [100], language is a like a boat on the sea of life. Our cultural ancestors have created more and more comfortable boats to help us swim, so that we have lost the courage to jump into the water. But in order to understand our own methods of thought, we must start without a boat at our disposal and try to reconstruct it plank by plank. This is possible because we can count on our "swimming abilities", i.e., we can rely on our language capacities without relying on a particular language [97]. The operations for referencing and predication proposed here can be used not only to reconstruct natural languages, but also synthetic and formal ones, in which information ontologies and data models may be expressed.

Note that although this view remains speculative, it is supported by many findings in attentional studies (Section 3.2) and can be exemplified by concrete examples of perceptual operations (see Section 5.1). It is furthermore general enough not to make any assumptions about how these operations are actually realized, and therefore remains open for any cognitive model that complies with the general assumptions made so far. The theoretical framework outlined here will be used as an informal background theory in the remainder.

Changing perspective and moving attention. Mental domains are collections of things human attention can be *focused on*. If we take entities into focus and store these discrete moments of attention in memory, we call them *foci of attention*.

The process of continuously paying attention to a given domain and storing the generated foci of attention is called *scanning*. Scanning always induces a *temporal order* among the foci of attention (processing time), and it also records structural relations. Mental domains can be taken into different perspectives by *attentional scope*. The attentional scope constrains human attention to some subset of the domain, and may be thought of as a moving window in that domain. While attention can focus only at one (or at most few) different things, the attentional scope may encompass a whole domain.

Intermediary spatial reconstruction (ISR) and referencing. There is one basic mental domain, which I call the *intermediary spatial reconstruction* (ISR) (Section 3.2), which is special because it enables to learn referencing and helps grounding other domains of human imagination. It is unconsciously constructed in the brain from direct sensory-motor inputs, integrates different perceptual modalities (vision, palpability, hearing) into a coherent view on a perceived scene, and accounts for the fact that perceived reality appears volumetric and continuous from a certain point of view even though it is a reconstruction from 2-D signals (inverse optics problem). It corresponds to Tversky's *space around the body* [191] and Montello's *vista space* [117], and includes the observer himself as he or she walks about the environment (compare Figure 3.5). In the remainder, I will sometimes use the term *space around the body* to refer to the ISR domain. This is in order to distinguish it from other domains of spatial cognition. Note also that this space is an unconscious mental product of a human observer and not something that exists independently, even though it is our primary window to the world.

The ISR domain is basic in the sense that it helps to (re)construct other domains that the human mind can imagine. Other domains include, for example, the *space of navigation* [191] or environmental space [117], associated with cognitive maps, geographic cognition, landmarks, and way finding. This domain needs to be constructed from memorizing entities experienced in the ISR domain while moving about the environment, or from viewing maps [88]. Furthermore, there is the domain of *conceived time*, which includes filled in time intervals that have never been experienced directly (e.g. the day of death of Napoleon), but which is partially reconstructed from processing time. And there are *abstract domains*, which include entities without spatio temporal unity, normally referred to by general terms or universals (categories, qualia) and the like.

The act of *referencing* is learned as an act of attentional focusing which *associates* an entity in focus with some name, definite description or deictic symbol. Through the mechanism of *joint attention*, referencing can be *shared and calibrated* across human minds. The mechanism of joint attention *is only available for the ISR domain*. This is one reason for its fundamental importance: shared referencing of entities in other domains (if they are not mathematical, see Section 3.3.2) *needs to be accomplished via ISR, either directly or indirectly.*

Predication as scanning Gestalts. The part of a mental domain which is in attentional scope is a subset to which *Gestalt mechanisms* can be applied. These are unconscious pre-conceptual mechanisms which self-distribute over a domain and produce stable structures called *Gestalts*. They often appear as bounded regions, but may also appear as fields of force, structural equivalences, or continuous fields, as in the case of color (Section 3.2). Some of them may be triggered automatically by a scene, some of them may be consciously applied and thereby influenced by human reason. Many of them are the result of learning, in the sense of internalized sensori-motor patterns, as proposed by Barsalou [5]. In this way, conscious human reasoning can be internalized, and therefore is reflected by bodily mechanisms[15]. Once it is available, a Gestalt mechanism's functionality itself is not based on conscious reasoning, and can thus account for the necessary grounding and tractability of perceptual knowledge (Section 3.2).

The presence of a Gestalt in a domain is recognized by the focus of attention[16]. A *predication* is an act of *scanning* a mental domain *and recording* the presence of a Gestalt at each focus of attention. In this way, the Gestalt and its form enters conscious experience. Predication requires prior (conscious or unconscious) application of a Gestalt mechanism. The results of predications are stored as n-ary relations on memorized foci of attention. Each such relation therefore can be associated with a unique Gestalt mechanism, which allows to distinguish them even if they may coincide.

If predications are performed on the ISR domain, and therefore invoke Gestalts in the domain of immediate spatial perception, I call them *perceptual operations*. Because this domain is so basic, its Gestalt mechanisms are of particular importance. Possible Gestalt mechanisms as a basis for perceptual operations are discussed in Section 5.1, including ones for individuating bodies, actions and other processes.

Learning *predications* in language involves associating symbols with a scanning operation, as argued by Langacker [92]. But note that it may also involve learning a new kind of Gestalt mechanism, since humans are not equipped with all of them from birth ('boundary introduction').

Constructions. In the course of mental construction, we usually begin with paying attention to the intermediary spatial representation ISR, and with performing predications. Other domains are constructed using *reification operations*, based on some underlying perceptual relation, the scope of attention, and some *criterion of unification*. Possible ways of construction result from combining these factors. For example, *mass-noun construction* can be based on attentional scope. *Count-noun construction* may use a unity criterion with respect to a perceptual

[15]Compare Langacker's [92], Johnson's [74] and Lakoff's [90] ideas about how logic and inference may be grounded in body based Gestalts.

[16]This is not to claim that we "see" Gestalts consciously, rather that we recognize entities in the world through them.

relation (as illustrated in Section 3.3.3.2), for example a maximally self-connected class of foci of attention.

It should be noted that I expect not every domain to be constructible in this way. In fact, I suspect that many mental domains are results of pure human imagination [74], being *mere postulates* without much import of experience[17]. As argued in Section 2.2.1, Quine's idea of the network of knowledge loosely coupled with experience [136] is a case in favor of this view. Possible operations for construction are discussed in Section 4.1 against the background of relevant philosophical and logical literature.

I admit that the proposed view is largely influenced by indirect realism [134], cognitivism [91] and constructivism [195, 100], and that it owes much to the moderate empiricism of Quine [137]. But it is in principle neutral with respect to any metaphysical view. What it distinguishes from available cognitivist theories is its focus on reproducible operations and inter-subjectivity. Any theory about semantic grounding must primarily be able to explain how humans actually accomplish inter-subjective measurement and observation, despite all the cognitive and linguistic ambiguities involved. What it distinguishes from available constructivist approaches is its focus on perceptual and attentional mechanisms. Nevertheless, the view proposed is neutral regarding the relation of knowledge and reality, and thus does not exclude a moderate realism. Also, its constructivist solution makes it open to account for the creativity and diversity of human perspectives on issues of ontology, which seems an important requirement for any semantic theory.

[17] Compare Carnap's idea of theoretical postulates described in [24].

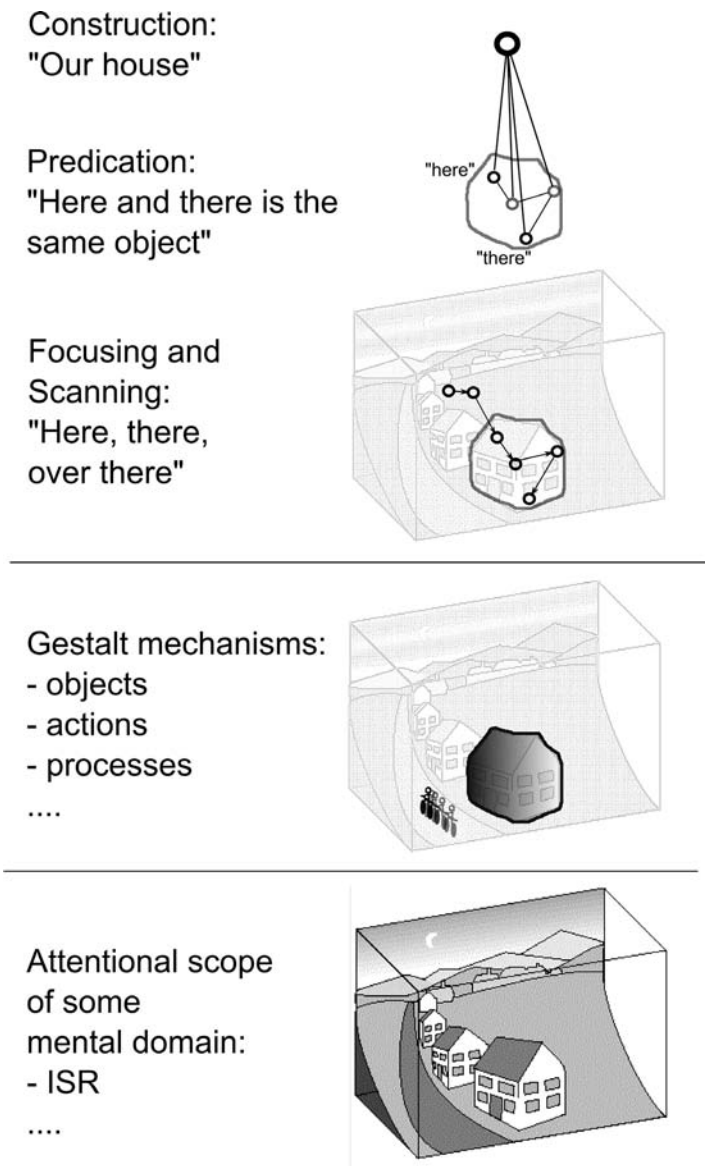


Figure 3.14: An operational view on predication and referencing. Panorama adopted from Steven Lehar ©(by kind permission), cf. [94].

Part II

Sources and methods for data grounding

Chapter 4

Constructive sources for data grounding

Our talk of external things, our very notion of things,
is just a conceptual apparatus that helps us
to foresee and control the triggering of our sensory receptors
in the light of previous triggerings of our sensory receptors.
The triggering, first and last, is all we have to go on.

— W.V.O. Quine [141]

Information ontologies [56] are a formal kind of meta-data. They allow to describe data, to infer data, and constrain its interpretation [84]. Like meta-data, data consists of formal symbols. More specifically, it can be regarded as a collection of *ground sentences* in some formal technical language. Ground sentences are logical formulae without any variables, i.e., they contain only predicates and names, and thus express facts about individuals. Information ontologies can be regarded as logical theories in this very same technical language. In contrary to data, they contain variables and quantifiers, and thus can express facts over sets of individuals. From a constructive perspective, the grounding problem of information science (Chapter 1) requires us to reconstruct this technical language based on operations of predication and referencing [100].

One of the main problems of information ontologies lies in their choice of what to consider as primitive, and what to consider as derived. At which level of abstraction should we belay our ontology? If we follow Quine's thought cited above, this problem is not surprising. Our choice depends as much on the usefulness of a conceptual apparatus in the light of experience, as on our experience itself. A similar view was taken early on by Hans Vaihinger in his *Philosophy as if*[1]. What distinguishes my approach from Quine's is that I do not see experience as triggered by sensory receptors, but rather triggered by Gestalt mechanisms.

[1] "Die Begriffe sind als Durchgangspunkt gleichsam die Scharniere, durch welche die Verbindung der Empfindungen hergestellt wird", [197, Chapter 25].

Information is therefore a category very different from data as well as metadata. Sharing information involves the whole pragmatic apparatus of reference described in the last chapter. It means that a data user has learned to perform the same predications, and is able to follow the attention of the data provider, including the various ways of abstraction involved. A focus of attention, accordingly, can be seen as a *lowest level information item*, since it can be shared by communicative agents. In the following, I suggest a way to formally construct *abstract information items* based on (1) joining attention on reproducible experience and (2) following constructive procedures. Both kinds of operations can be guided by a formal technical language. In this chapter, I will make suggestions how formal primitives for predications can be established in terms of an *observation language*, and how this language can be used to *guide the construction of abstract things* we refer to in ontologies or data sets. Observation languages are communicative conventions which involve learning and teaching predications and referencing. This competence builds on perceptual operations (Section 3.4), and thus lies partially outside the scope of any symbol system, including this thesis. What is in focus in this chapter are the *constructive means* provided by such a language once it is established, keeping in mind that they only insufficiently mirror those *mental constructions* that underlie human intelligence[2].

I begin in Section 4.1 with arguing that the arbitrariness of ontological abstraction is partly a consequence of the presence of language fictions [12], and that the underlying constructive freedom needs to be taken seriously. I also argue that it is feasible to guide the construction of fictions using formal observation languages, established in terms of the referential apparatus of Section 3.4. In Section 4.2, I will introduce *constructive language tools* in a grounded first-order logic, based on observation predicates, existential quantification and individuation criteria. Many suggestions I make are inspired by Quine, as well as Vaihinger's treatment of fictions in [197], but some need to remain only sketched.

The suggestions made are *constructive sources* for a *meta-theory of grounding*, i.e., for a theory that allows to build grounded theories, and thus semantic reference systems, in a community of speakers.

4.1. The construction of language referents

In order to account for the obvious human creativity in thought, we must assume constructive operations that allow to generate ideas about the world. Practical knowledge about these operations is needed, because it is their results to which humans refer to with language. In the remainder, I will first argue that denying this constructive freedom causes a problem of arbitrariness in current information ontologies: it is the arbitrariness of Bentham's useful language fictions or Quine's

[2]The latter seem to rely on hierarchical auto-associative simulations [64] or other Gestalt-like mechanisms.

logical reifications we are confronted with[3]. Then I will discuss ideas that help address the underlying operations. I will first discuss constructivist suggestions made by Piaget, Glasersfeld and Lorenzen. These authors consider abstractions as results of operations performed on representations of other mental operations. Then I will discuss how this construction process can be guided by a formal observation language, based on ideas of Carnap and Quine. I will first critically review their philosophy, before I make a suggestion of what can be retained and defended in the light of contemporary critique. How an observation language can be established in terms of the apparatus of Section 3.4, and how constructions can be guided with it, is explained subsequently.

4.1.1. The arbitrariness of ontological abstractions

The *Theory of Fictions* [12], published in 1815 by Jeremy Bentham, is a noteworthy contribution to the problem of language semantics, which inspired Ogden in his treatment of the meaning triangle [122]. Among other things, Bentham states that individual percepts, such as those of bodies, can be used to expose the meaning of *language fictions*. Language fictions are linguistic creations of the mind which cannot be directly observed, but depend on other perceivable entities. Bentham's treatment of language is a treatment of names, i.e., nouns and their compounds which denote particular entities in thought. For instance, the sentence *the color of this body is red* may denote three mental entities, a body, its quality, and a thing called redness. According to Bentham, such entities can either be *"real"* or *"fictitious"* [12].

"Real" entities are the ones whose existence is non-disputable by the speakers of a language. This class mainly consists of *perceptibles*, i.e., individuals that can be identified via perception. These include *bodies* (physical entities) in the first place, as well as *individual percepts* like *pain* and *pleasure*.

In contrast, "fictions" are posited by a speaker because they are useful for communication. Fictions serve as language proxies for real entities and often depend on their existence. For example, redness depends on the existence of a concrete color percept and on the perceived part of the surface of a body. Without those it would not be sensible to talk about redness as an existing entity. For Bentham, reifications of quality values are therefore fictions in the same sense as the entities denoted by the nouns *motion, cause, action* or the legal fictions *obligation* and *right*. All these existentially depend on other directly perceivable entities [12].

What is remarkable is that Bentham, in contrast to many nominalists or empiricists, acknowledges the existence and usefulness of fictions as abstract entities, while making clear that such abstractions are a consequence of constructive linguistic freedom and therefore need exposure in the light of observable roots.

[3]The following subsection is a short version of ideas first presented in [159].

I argue that Bentham's fictions are what modern logicians such as Quine [138] would have called *reified logical entities*. Furthermore, Quine rightly alluded to the fact that such reifications can change the underlying *ontological commitment* of a theory [138], i.e., the commitment to what exists or does not exist. Quine proposes in [138] that reified entities can simply be conceived as individual values of existentially quantified variables[4], and therefore need to exist in the domain of interpretation of a theory.

Similar to Bentham's suggestions how to sidestep fictions[5], Quine sums up that "whatever we say with the help of names can be said in a language which shuns names altogether"[6]. Thus, unnecessary (fictive) names can be avoided by introducing predicates which rephrase them as descriptions. For instance, to avoid talking of a fictive particular *Pegasus* as an existing entity, we can refer to it as *the winged horse captured by Bellerophon*. And to avoid talking of *classes or universals* as existing things, e.g. 'humanity', we can introduce a predicate like 'human'. Bentham similarly reckoned all sorts of *aggregations* (whether collections or classes) among the fictions [12].

Quine concludes that these *reifications are arbitrary* to a large extent [138]. This arbitrariness can be considered a manifestation of the empirical indeterminacy of theories, and Quine's idea of an empirically underconstrained network of knowledge (compare Section 2.2.1). Carnap [23] pointed out earlier that the practical acceptance of reified entities in science or natural language is a matter of *practical commitment* to a linguistic framework which requires them. It is not a matter of committing *on principle* to some metaphysical view, as the one of realism or nominalism. For example, the choice of real numbers instead of rationals to represent physical space, and thus the existence of irrational numbers in physics, is a matter of practicality (since Euclidean geometry requires irrationals). It is not a consequence of measurable facts about existing things, because any quantitative measurement yields at most a rational number. According to Carnap, the question of metaphysical existence of entities is irrelevant and not a question science should be concerned with. In fact, this renders any a-priori argument in favor or against the existence of abstract entities useless. It also shows that philosophical discussions about the existence of universals, continuing to the present day (compare the summary in [2]), are concerned with a pseudo problem.

What makes the arbitrariness a *practical problem of ontology engineering*, is that any reification or deconstruction of the sort described above is a substantial change of the theory, since the domain of discourse expands or shrinks. Consequently, such degrees of freedom in modeling the real world tend to produce

[4]This seems the root of his many cited and misunderstood claim "to be is to be the value of a variable".

[5]Compare in Bentham's book [12, pages 86 ff.].

[6]We remain committed to the existence of entities "at least until we devise some way of so paraphrasing the statement as to show that the seeming reference to species on the part of our bound variable was an avoidable manner of speaking" [138].

non-interoperable ontologies. There cannot be one-to-one correspondences between semantic domains if the underlying ontological commitments differ.

Contemporary information ontologies are affected by this arbitrariness of reification, which is one reason for incompatible top level ontologies. For instance, Neuhaus et al. [118] argue for the existence of universal entities and reified qualities such as 'universal sphericity'. They introduce generic dependence relations among particulars and their instantiated universals which exactly match with Bentham's ideas of existential dependence. Additionally, they reify a layer of quality fictions orthogonal to this one, building an 'ontological square': quality particulars inhere in other particulars, like in *the particular redness of this apple*, and instantiate quality universals like *universal redness*. Other ontologies such as DOLCE [109], use qualities as individual entities that inhere in only one object, e.g. every physical object has a particular color quality. However, the authors exclude reified universals like *appleness* in their domain of discourse. In another paper, they demonstrate that there are many possible more or less reified formalizations of an ontology of qualities with roughly equal expressiveness [109]. Grouping quality values in value ranges such as "red" results in abstract color qualia. Probst has demonstrated that such reifications can be necessary in order to account for the complex dimensional relationships between objects and their qualities [132]. Bateman thus rightly argues [7] for a flexible *layered ontological framework*, which can account for the granularities of different views on space.

Most such ontological variations are useful and plausible by themselves. But it is their arbitrariness which makes it hard to reach semantic agreement. In their attempt to construct a cognitively plausible theory, the cited authors in a way just illustrated that language and thought are indeterminate, creative, and full of (useful) fictions. But this is the reason why semantic ambiguities exist in the first place.

Since introducing abstract entities in an ontology is a matter of choice as much as internal coherence with experience, I suggest to take this freedom seriously instead of arguing against it. A theory of information grounding therefore needs to be not only about the inter-subjective perceptual apparatus for grounding, but also about the concrete ways of constructing abstract entities in thought. It needs to be, in some sense, a *meta-theory about constructing empirical theories*.

4.1.2. Abstraction means to reflect on represented actions

As a primary result of his cognitive studies, Piaget [126] suggested that abstract ideas, concepts and categories are a product of *reflection*. His basic insight is that cognitive agents do not abstract from sensory-motor inputs by applying knowledge, but by reflecting on their own mental actions. Let us illustrate this idea by the concept of natural numbers, whose construction was imagined by Piaget in [126] in the following way. I made Piaget's ideas more concise by drawing on ideas of Lorenzen [99].

In order to understand the *cardinal numbers*, children need to draw on counting operations. These consist of the capabilities to (0) *identify* individuals and to (1) *classify or aggregate* them in a perceived scene, as well as to (2) *induce an arbitrary order* on them by jumping from one to another with their focus of attention ("counting")[7]. The output of this operational pattern, which we could call a *counted class pattern*, allows to construct equivalence classes by *one-to-one correspondence*. In each counted class, the first one and each follower entity are recognized based on their counting order in that class, not by any perceivable qualities. This independence from perception is essential in order to construct mathematical abstractions, like numbers. An equivalence relation among two counted classes holds if and only if each entity in one class has a corresponding one – with respect to its position in the counting order – in the other class. The respective equivalence classes among counted classes are cardinal numbers, however only a very incomplete subset of them.

Note that in each constructive step, we had to draw on memorized results of certain mental operations. What we can see from this example is that the roots of formal conceptual structures are neither a result of perceived qualities nor a pure formalist invention, but rather a matter of abstraction from certain operations. In this view, humans, like all animals, are equipped with certain attentional capabilities to act on sensory input. But it is their ability to consciously act upon *representations of such actions* (in this case memorized results) which allows them to build abstractions. This constructivist claim (compare [195], Chapter 5) needs deeper inspection.

Actions and reflective abstraction. What exactly are actions or operations? *Actions*, in Piaget's sense [126], are repeatable, intentional events. *Operations* are special kinds of actions, namely ones that are interiorised and reversible. This means they can be simulated in your mind in the absence of their perceptual context, and can also be reversed, performed from outputs to inputs [126]. Actions can be coordinated with each other, resulting in new actions (called *coordinations*). For example, they can be combined in parallel or in sequence, or they can be composed, such that the results of one form the input to another one. But in order to abstract from concrete actions, they need to be represented on a conscious level.

Actions may transform external (i.e., perceived) states of a scene. For example, putting a cup on the desk. More importantly, they also transform internal states of the mind of an observer, e.g., a state of memory [195]. Only if actions are reflected in states of memory, they can be remembered at all, and thus, be represented in mind. And only if they are explicitly represented, we can consciously pay attention to them[8]. The conscious abstraction process was called

[7]Compare second lecture in [126]. I added capability (0), which is only presupposed by Piaget.

[8]Very often, we do not remember or represent actions in mind. For example, we can play

reflective abstraction by Glasersfeld [195, Chapter 5] and Piaget. Note that consciousness implies the workings of the attentional apparatus described in Section 3.4. I suggest therefore that reflective abstraction implies paying attention to action representations.

How can actions and operations be represented in mind? We have to distinguish *action types*, i.e. dispositions to act, from *particular action performances*, or *particular results*. All of these can be input to abstractions. Actions may be represented therefore

- by *their structured results stored in memory*, e.g. their inputs and outputs, or
- by *single entities* standing for actual performances or operational schemes (compare Glasersfeld's operational consciousness [195]), or
- by *recursive rules* (standing for operation types), as Lorenzen argued [98].

In the first case, we can remember inputs or results, in the second case we may not be aware of any actual performances. In the last case, however, we represent an operation by specifying how results may be generated, potentially ad infinitum. *Abstractions* are built by operating on these kinds of representations.

If we combine Quine's insight about abstraction as logical reification with these ideas, we can conceive abstractions as *logical reifications that reflect action representations in direct or more indirect ways*. In the case of natural number construction illustrated above, the ordering relation in memory is a direct output of the counting operation, and each counted class is the output of an aggregation. Suppose these outputs are stored as relations in memory. The relations can then be the input of another operation that produces equivalence relations among the counted classes. Another operation may be used to build corresponding equivalence classes based on them. These classes are *extensional logical reifications* that stand for natural numbers. They directly derive from the underlying results, and thus have an extension in memorized experience.

But this does not answer yet how to close the cardinal numbers, and how to account for *mathematical fictions* like infinite sets. The natural numbers are infinite. We do not even construct finite but large numbers, say, 1.000.000, bottom up. I suggest that this kind of abstraction simply requires another kind of *logical reification by analogy, which may be based on a recursive rule representation*. As Lorenzen suggested in [99, 100], we may represent the counting operation in terms of recursive rules in a calculus of the natural numbers:

$$
\begin{array}{lll}
K_{nat}: & & \\
 & \longmapsto 1 & (\textit{beginning}) \\
 & x \longmapsto x1 & (R)
\end{array}
$$

tennis without the slightest attention to the motor operations involved, even though they were once consciously learned.

Using R, any natural number, e.g 11, 111, and so forth, may be mentally generated by substituting its predecessor for the variable x. While this is of course an incomplete representation of counting, it allows to generate fictive countings in mind without caring about the underlying perceptual operations. It also allows to introduce the transfinite cardinal number $\aleph_0$ (the cardinality of the natural numbers) as a logical reification based on K_{nat}.

Formal languages about human operations. It is clear that the result of number abstractions can be expressed in formal theories like Peano arithmetic. But how can operational semantics of the kind described here be expressed in formal languages? This question has by no means a straightforward answer.

Operational outputs may be described by relations. Relations can be established in memory a posteriori among any constellation of remembered foci of attention. In this way, one-to-one correspondences between classes counted in advance can be established afterwards.

In this thesis, human operations are expressed by their relational outputs in memory (if we need to talk about performances) or by some entity standing for the type of operation. The output is most easily captured in language by what I will call an *observation predicate* (see next subsection).

Another approach would be to use a recursive rule language such as Lorenzen did in his operative mathematics [98]. I have used this idea in a recent paper in order to give the experiential geometry axiomatized in Chapter 6 an operative "Aufbau" [155].

But we have to accept that there is some point beyond which definite language descriptions of human actions fail. In this sense operations can at most be *guided or recorded* by formal languages, but they are not substitutable by them.

4.1.3. Observation languages and quantification as constructive tools

Language can be used not only as a tool to communicate experiences, for example to predicate qualities of objects. It is also a primary tool to guide the construction of domains of thought from experience. In this subsection, I will look at how language philosophers have imagined the role of formal language in this context. I will focus on empiricist (but non-nominalist) formalists, such as Rudolf Carnap and W.V.O Quine. Both had useful ideas about conceptual construction from an experiential basis using the tool of formal languages. Even though they largely influenced modern analytic philosophy, some of their empiricist and formalist standpoints were rightly criticized. I will position my proposed solution with respect to those ideas and criticisms.

Carnap's formal reconstruction of science based on experience. The idea of formal reconstruction based on experience already underlies Carnap's most influential work *Der logische Aufbau der Welt* [25]. Carnap basically proposed a partially

formalized logical theory, called 'Konstitutionssystem'. It was intended to contain all notions of science and common sense as part of its vocabulary, constituting them from experience primitives. A notion is constituted in this system, if all sentences about it can be transformed into sentences containing solely experiential primitives. This basically requires notions to be either explicitly definable on the basis of such primitives, or indirectly as *definition in use*[9]. Beginning with elementary experiences ('Elementarerlebnisse') of a single individual, called the 'autopsychological' level, Carnap proceeded to the physical, and then to the 'heteropsychological' level constituting other human beings, and ended with the social and mental.

Carnap's proposal implies the fairly radical assumption of *empiricist reduction*[10]. He assumes that all sentences in science, expressed in this system, can be unambiguously transformed into a sentence in which only directly observable primitives occur as nonlogical constants. This claim was famously rejected by Quine in terms of the 'second dogma' [136]. He argued

1. that empirical theories are generally underconstrained by experience. So, different but incompatible theories can account for the same experiential evidence. Sentences in one theory may not exist in another one or may relate differently to other sentences. Therefore singular statements (as well as the entities posited in them) "face the tribunal of sense experience [...] not individually but only as a corporate body" [136]. A sentence thus can never be related to experience in isolation.
2. that theories are indetermined regarding their interpretation into experience. This means that even a single formal theory fails to relate unambiguously to experiential evidence. Quine shows in [136] that Carnap's project fails already when he tries to assign perceptual qualities to abstract space-time points in a reference system by an underconstrained system of rules [25, §125]. This point also makes it impossible to distinguish a-priori between observation terms and theoretical terms across human beings, as terms may be interpreted differently[11].

I suggest that in the face of this critique, one should give up the idea of a single rational reconstruction, since there may be several admissible ones, as argued in Section 4.1.1. Also, the establishment of observation languages is a matter of convention, no deterministic enterprise. Furthermore, we should be aware that even though humans are able to reconstruct entities in thought, they operate with an extra-linguistic experiential machinery which cannot be dispensed in favor of

[9]Carnap's ideas about definition in use contain some useful constructive suggestions exploited in Section 4.2.

[10]Carnap later suggested to get rid of the reductionist assumption and to allow for more loose kinds of term introduction, but to stick to the idea of *objective rational reconstruction* [24].

[11]For the same reason, the notion of 'analytic' truth, the first of Quine's criticized dogmas, remains a problem.

a pure formalistic approach. This contrasts with the logical positivist assumption that science consists of pure structural descriptions [25, §16].

Quine's observation sentences and the roots of reference. Quine worked on the project of an empiricism without these dogmas for the rest of his life. His proposed solution for the problem of inter-subjectivity, as I conceive it, is based on the notion of *observation sentences*. These are occasion sentences like "there is smoke" or "there is Mama", to which a community of speakers consistently consents to in a jointly observed situation. These sentences build an entrance gate to language, because they can be learned easily by ostension based on shared perceptual operations and joint attention in a group of people. Observation sentences in this sense *correspond to predications in the space around the body (ISR)*, as described in Section 3.4, and are part of the mechanism for learning reference. They closely resemble predications in constructivist language games as proposed by Lorenz [96] and Lorenzen [100, 76]. Eco's 'semiotic primitives' can likewise be understood in this way [35].

Before I explain Quine's ideas about constructing empirical theories, I will try to defend this idea against some of its critics, and also against some of Quine's own obscurities.

Quine's idea, stated in the form above, avoids many philosophical traps: First, it avoids being captured in language like formalist theories of language meaning, since it relies on the extra-linguistic operation of joint attention. Second, it avoids objectivism, since it does not require humans to refer to objects without their senses. Third, it avoids positing language independent facts of experience like in traditional empiricism, since observation sentences bear terms reusable in other parts of a theory. Fourth, it avoids relativism and skepticism, because it involves a practical solution of the problem of inter-subjective language reference, as explained in Section 3.3.3.1. And lastly, it avoids single perspective and reductionist accounts of science, as in Carnap's approach.

The idea was attacked by Donald Davidson and other objectivists, partly due to Quine's own obscurity on the matter. The problem is that Quine has formulated many more or less objectivist versions of it. In his most famous book *Word and object* [143, §8], he tries to objectify the meaning of observation sentences by the idea of *stimulus meaning*. This seems to cause unnecessary problems. Since Quine rejects to look into an observer's head (due to his own behaviouristic dogma), he needs to identify the meaning of an observation sentence with the class of irradiation patterns that reach the observer's retina and prompt his or her assent to it. Davidson has rightly argued in [30], that this idea falls victim of the problem of inter-subjectivity. The reason is that *a stimulus at an observer's retina cannot be shared among observers*. Stimulus meanings therefore fail to deliver the required explanation of how humans can share object identifiers in a scene perceived from different angles, and how they know that others are looking at the same scene. What is necessary to solve this puzzle is the apparatus of joint

attention and a pre-conceptual basis of comparison[12], which were outlined in Sections 3.3.3.1 and 3.4. What is *not* necessary, is Davidson's objectivist account of meaning, that tries to do without perception.

Since observation sentences are occasion sentences, their truth only depends on the observed situation. How can they relate to "eternal" sentences of a theory, whose truth is not situated? And how does this lead to the construction of abstract entities one can refer to? The *empirical content* of a theory is posited in terms of *observation categorials* [140]. Observation categorials express conditioned expectations. They are experiential implications of the form $\phi \to \psi$, where ϕ and ψ are two observation sentences whose deixis point to the same observable spot, as for example "Where there is smoke, there is fire" [140]. Note that these rules are *inductively posited*, because they can never be derived or verified from observation sentences. They are part of Lorenzen's *predicator rules* that interrelate predications in a language terminology [100] and further standardize its use. But note that they can still be used without any appeal to the construction of individual entities in thought. Observation categorials can just be learned in terms of joint predications in a situation. In contrast, "'eternal" sentences of a scientific theory require the construction of entities, because they entail observation categorials, which means we need to substitute nameless individuals for terms. This requires *quantification over referents* for the first time [137]. Quantification is an unambiguous sign of mental construction of singular entities, because it *presupposes individuals* to quantify over[13].

The process of quantification in conjunction with observation sentences is what Quine calls the *roots of reference* [137, §26]. It can be regarded as a basic language tool that indicates and guides mental constructions.

4.1.4. Ways of constructing domains of reference with logic

I will suggest in this subsection how first-order logic, as a technical language, can be used to guide the construction of domains of reference. One should not confuse language as a tool with the actual construction process, even though it is its most essential manifestation[14]. As I have argued in Section 3.3.2, the process

[12]In principle, this is also required by Quine when he speaks of a "pre-linguistic quality space" in [143, §17].

[13]For the theory to be generally applicable, it needs to cope with *nameless entities*. For example, it needs to cope with an unbounded realm of unknown electrons, persons and other individuals beyond the ones immediately known to observers. But in order to be applicable to concrete observations, nameless entities need to be *inserted* into observation categorials. For example, the theory may predict the existence (via existential quantification) of some nameless electron e satisfying some condition ϕ. Only if it is possible to insert this electron into a variable e in "If e satisfies ϕ, then a particle detector observes it", then an observation sentence of the form "particle detector observed e" can be derived.

[14]This distinction seems not to be clearly drawn by constructivists like Lorenzen [100]. It is also the reason why the commitment to classical logic does not prevent one from taking a

of constructing mental domains remains hidden from direct observation. Quine argues that it is *partially reflected in language*, but that it is still obscure to a large degree[15]. How entities might be constructed and referenced may be learned from looking at language acquisition. A possible account of this ontogeny was pinned down in [137] and [141]. However construction may actually be performed: For referencing purposes, we need to be able to guide it by a language such as the one proposed in the following.

Constructing observation predicates from predications. How can observation sentences, which in our view correspond to individual predications (i.e., actions) in the ISR domain, be used to construct *formal observation predicates* (called 'general terms' by Quine)? These are indispensable for quantification [137]. Quine proposed that predicates may be extracted from observation sentences such as "this is white", by *relative clauses with relative pronouns*. These are of the form "thing x, such that x is white" [137]. The pronoun x plays the role of a variable which can be applied to more than one term[16]. Morover, *polyadic* predicates such as "part of", "darker than", "bigger than" can be extracted in a similar way. In terms of our operational semantics, the application of such a predicate to some term "a", like "a is white", simply denotes another predication in the ISR domain.

I propose therefore that the *values of variables in observation sentences are bound to foci of attention*, which are memorized moments in which a human observer focused on the ISR domain. At the same time, observation sentences are records of speech acts. In consequence, observation predicates are *extensional* in the sense of first-order logic, i.e., in the sense that they have the usual Tarskian interpretation in terms of sets of foci. Note that this should not be confused with their meaning, which is given in terms of predications, i.e., actions[17].

Combining predications with truth functions. Predicate applications can be combined by the usual logical connectives or truth functions, e.g., "x is white and x is higher than Mont Blanc", in order to define more complex observation sentences. It is also possible to define more complex predicates this way[18]. The

constructivist viewpoint.

[15]"A common man's ontology [...] is vague in its scope; we cannot even tell which [...] things to count him as assuming. Should we regard grammar as decisive? Does every noun demand some array of denotata? Surely not; The nominalization of verbs is often a mere stylistic variation." [141]

[16]Similarly, names or 'singular terms', like "x is Fido" are applicable to only one term.

[17]Note that the extensions of observation predicates constantly need to change as the observer continues to experience the world. Predications relate moments of attention by a Gestalt present in the ISR. But attentional moments enter memory and fade from it again in the flow of conscious experience, while the predicate continues to exist. Furthermore, two different predications may be performed on the same occasions, as Gestalts can overlap. Observation predicates therefore are semantically bound to the underlying Gestalt mechanism, not to the set of attentional moments stored in a relation.

[18]For this last purpose, I assume the well known λ operator.

extension of combined sentences thereby corresponds to the usual Tarskian algebraic combination of predicate extensions, i.e. of memorized foci of attention. From the viewpoint of operational semantics, negation of an observation sentence just means that the observer is unable to predicate a certain Gestalt in an observed situation. Conjunction means that two different ones can be predicated, and disjunction means at least one out of two[19]. For a more general constructive introduction of logical connectives, see Lorenzen [98].

Reifying entities by existential quantification. As Quine argues [138], existential quantification is not only a necessary ingredient of empirical theories, it also marks the point in language where mental construction of individual referents becomes inevitable. *We populate our ontology – basically – using the language tool of existential quantification*, i.e., by axioms of the form $\exists x.\phi(x)$, where ϕ is some predicate. Quine considers this reification an irreducible constructive step [137, §26]. In the simplest case, we may just posit a theoretical entity by an existential quantification axiom without any import to other entities in the domain of predication. In the more interesting cases of successful reference, however, we construct these entities relative to foci predicated over in observation sentences. But what kinds of construction principles do we use here?

Tying entities to ISR predications. Quine has proposed such a linguistically supported path of construction several times in [137, 141, 139].

We may begin by constructing dyadic observation predicates simply for those predications that focus on *body and texture Gestalts*. For example, we may follow and point at the body Gestalt of a person or look at the surface of a fluid and construct predicates like "x and y belong to Mama" and "x and y is milk", learned inter-subjectively in observation sentences. We may then define predicates which apply for the whole domain and range of attentional moments of the former ones, such as: "x PartofMama" and "y PartofMilk". At the next step, we introduce an abstract entity for each one of these predicates by an existentially quantified statement: "there is some entity that corresponds to PartofMama" (how exactly is a matter of individuation, discussed in the next paragraph). Then, we introduce a new name (a singular expression) "Mama" and "Milk" denoting this very entity. We now have populated our mental domain with one body and one quality value, have introduced names for them, and have tied them to perception. This allows us to point at Milk and Mama as individual entities. Using Gestalts in the ISR domain, we may similarly construct entities standing for category names like "Human" and "Dog". In this case, we conceive of "Human" as a name denoting a single reified entity, similar to what philosophers call a *universal* [2].

There remains much ambiguity in construction. "Human" may be conceived as a name of a single *universal* entity similar to "humanity", or it can be conceived as

[19]Note that logical connectives in natural languages in general may not behave that neatly, compare [137, §20].

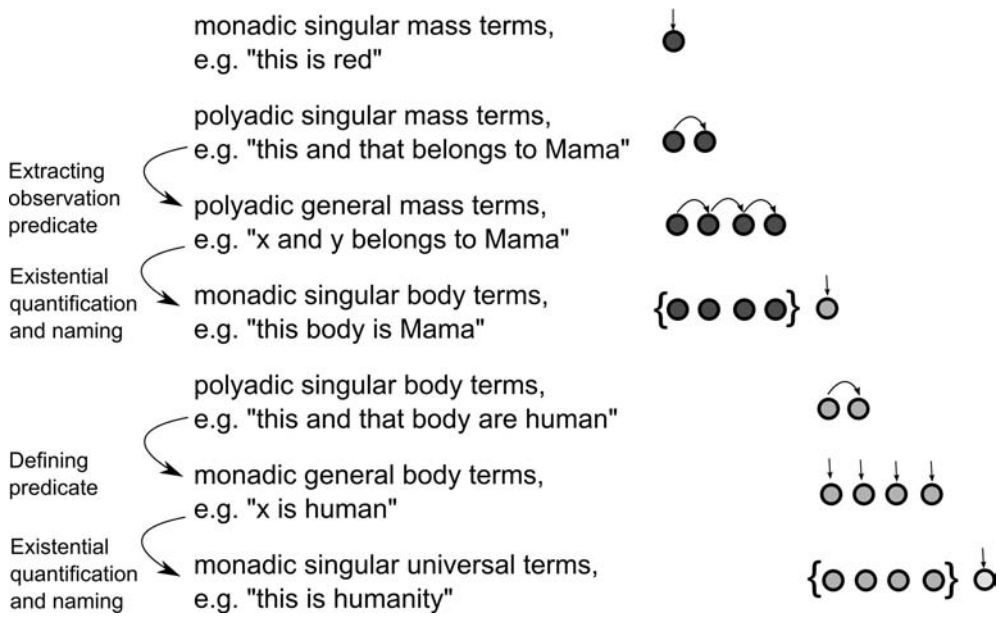


Figure 4.1: Quine's construction path for bodies and universals, cf. [137].

a predicate over concrete bodies. However, there is a way of explicitly referring to the entities these names or predicates are about, by formulating identity criteria for them.

Individuating entities based on observation predicates. The importance of *individuation criteria (IC)* for sound ontology engineering has been widely recognized [59]. They are in fact an essential source for exhibiting construction principles for individuals in ontologies. Perception is the key to distinguish and recognize, i.e., to identify, entities. In some sense, this means that it accounts for how the identity and tractability of experiential entities come into being.

Individuation requires criteria of *unity* (i.e., for constructing integral wholes out of related parts) and *identity* (i.e., allowing to track entities and distinguish them from each other) [59]. Even though these two aspects are often distinguished, as in [59], it seems plausible to consider *identity as construed based on unity*, along the following lines of thought.

To state *true identity (=)* of two things in an ontology is a very fundamental semantic constraint. It influences the resolution of the described world into indistinguishable chunks, i.e., the ontology's *granularity*. Quine claimed[20] that any two-place predicate expressing an *similarity relation* R could express identity relative to some sufficiently impoverished theory [139]. If, in this impoverished theory, two R - equivalent things become *indiscernible* with respect to the theory's terms (i.e., they satisfy Leibniz' criterion[21]), then the theory can always be

[20]See the *maxim of identification of indiscernibles* in [139].

[21]*Leibniz identity* is the criterion that requires indistinguishability of properties for identical

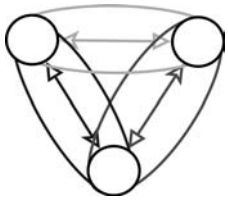


Figure 4.2: Imperfect community and companionship, cf. [54].

re-interpreted into a different (more abstract) domain such that R just means $=$, and equivalent names are *identified* with one and the same thing. For example, "if one has a language in which one speaks of persons and in which persons of the same income are indistinguishable, the predicates of the language may be reinterpreted so that the predicate which previously expressed having the same income comes now to express identity. The universe of discourse now consists of income groups, not people" [121]. This is Quine's proposal of how abstract things, e.g. 'classes', but also 'objects' and 'qualities', come into being [139]. At the same time it indicates how *identity criteria* can be constructed for them: An individual in the second interpretation, an income group, is an *equivalence class* of R in the first interpretation. So identical things in a more abstract (coarse-grained) theory can be expressed as classes[22] in terms of some similarity relation in a fine-grained theory. Williamson [204] called this mechanism of identification a *two-level criterion of identity*.

This actually leads us very close to Carnap's original approach, which was to reify abstract entities from primitives expressing perceptual similarity. Quine obviously was inspired by his friend and colleague Carnap here.

Possible objections. Such an enterprise is readily confronted with the standard objections to previous attempts of 'extensional abstraction of universals from resemblances' [2], like those in Carnap's *Aufbau* [25] (see also [127]). I will argue now why the approach suggested here is not affected by these objections.

1. Unlike Carnap's 'Ähnlichkeitserinnerung', this approach is *specific* and *modal*, since it is based on neuron states and perceptual capabilities. It is properly grounded in these capabilities, which are – as such – beyond question, even though their functionality is not explained yet. So philosophical puzzles about the proper specification of the resemblance relation do not arise. Also, there are several ones of them, one for each Gestalt mechanism.
2. General resemblance should be based on *particularized natures*, not vice versa (compare [2]). This is possible in my approach. A particularized na-

things in a theory [149].

[22]These were treated as maximal self-connected mereological sums, called *unities*, in ontological frameworks like in [59, 47], but the authors did not discuss identity criteria based on them.

ture of a thing comprises all the specific Gestalts present in focus, while a thing corresponds to a particular one of them. It therefore also allows to come up with *graded resemblance relations* among things based on partially intersecting particularized natures. For example, one "dog" Gestalt mechanism may be present for two animals, while their color and size differ.

3. The problem of *coextensionality of classes*, i.e., the problem of distinguishing between two properties that happen to inhere in exactly the same individuals. Suppose properties like "red" and "hot" were just coextensional classes. Then they could not be distinguished in a formal theory, since formal predicates denote nothing more than their extensions. This problem can be solved by observation predicates. An observation predicate denotes a perceptual relation which may involve a mental entity standing for the underlying operation itself. This distinguishes "hot" from "red". In this way, we make use of the fact that the identity of primitive observation predicates hinges on the underlying Gestalt mechanism, and not on the current set of entities stored in memory.
4. The problems of *imperfect community* and *companionship*, raised by Nelson Goodman [54]. The first objection goes back to Wittgenstein [207]. A 'family resemblance' is one where all members of a community resemble each other but in different ways, so that they may not have anything in common. For example, in the triadic community in Figure 4.2, each pair of entities has something different in common. Even though all members have nothing in common, they still would be treated as resembling each other in Carnap's approach. Furthermore, in his approach, perfect subcommunities that accompany the imperfect community, like each one of the pairs, cannot be distinguished. Thus a perceptual agent needs to know not only which entities are similar, but also *which are similar in the same respect*. As it turns out, this aspect is captured by particularized natures, see above.

4.2. Reifications expressed in grounded FOL

Individuation and other criteria for describing logical reifications may be formalized in a *grounded first-order theory*. In such a theory, mental operations are not described explicitly, e.g. as functions. Rather, the theory describes their results stored in memory of an observer. More correctly, the theory transcribes results of predication and construction actions taken by an observer in terms of an *observation record*. Such a record consist of tuples of foci of attention associated with various reified entities. Since reference to foci of attention can be shared among observers based on joining attention, and reifications can be described based on them, the domain of a grounded theory can be considered as consisting of lowest-level *information items*. This also means that a grounded theory does not mirror how cognitive constructions are actually implemented in the brain. It

rather describes how reference and construction can be carried out by an observer for communicative purposes. The following kinds of reifications are partially described by axioms and axiom schemata in FOL, and partially sketched informally. In the remainder, *predicates* are denoted by uppercase strings and *variables* by lowercase letters. All free variables in formulae are assumed to be universally quantified. Sometimes, we explicitly use *predicate wildcards* in axiom schemata indicating how axioms with a more specific purpose may be generated.

In Section 5.1, I will introduce a set of observation predicates. These are primitives of a grounded FOL theory. As discussed above, each of them denotes a type of predication in ISR whose result is stored as a relation on moments of attention. *The domain of moments of attention is finite and denoted by the unary predicate F.* F is an observation predicate and denotes a momentary window of attention on Gestalts in ISR stored in memory of an observer. This window changes in time, since old moments are deleted and new ones enter memory, and so does the extension of the predicate as well as the domain of the whole theory. Note that even though predicates of this theory therefore cannot be considered equal to their extensions over time[23], they can practically be treated so in a certain interpretation of this theory. Such an interpretation simply corresponds to a state of memory or a state of recordings, in which the extension is fixed[24]. This allows to employ the usual FOL semantics, in which non-logical symbols simply denote subsets on the domain of interpretation.

If an observer has recorded predications and attentional moments, one can perform abstraction computations on them. F denotes therefore a subdomain of the larger domain D of the grounded FOL theory, which may also contain reified "eternal" entities. The latter may roughly correspond to the observer's long-term memory, while F may correspond to short-term memory. Note that since reifications are actions taken by an observer, we cannot assume a reified entity to exist in our domain without further ado. Reified entities are not like classical sets. They need to be explicitly introduced by an existence claim. Since an observer can make only finitely many such claims, the domain D is assumed to be finite, too[25]. We can now define the notion of a reification simply as follows:

Definition 1 *Reification:*
$Rfc(x) \leftrightarrow \neg F(x)$

Note that I do not intend to provide a complete list of reification operations. Nor do I suggest that the diversity of human fictions, as sketched in Vaihinger's

[23]Compare the discussion in the last Section 4.1.4.

[24]In order to model the evolution of memory, one would have to distinguish different interpretations of the theory as temporal states of memory, or even model a sequence of theories.

[25]General finiteness cannot be expressed in FOL as a consequence of Gödel's *compactness theorem*: Every set of FOL formulas with arbitrary large finite models also has an infinite model. Thus finiteness of D is therefore implicitly assumed in the following.

opus magnum [197], could be comprehensively captured in this language. My goal is rather to devise some practically useful information construction tools.

4.2.1. Extensional reifications

As was discussed in the previous section, logical reifications can depend more or less heavily on the existence of other entities in the domain of discourse. One way of tying reifications to foci is to individuate them as *classes of foci*. One needs to be aware of the fact that many kinds of information, e.g., predictive inferences or information that transgresses the resolution of human attention, cannot be modeled in terms of classes of foci. Furthermore, there are philosophical arguments against set-theoretic approaches towards cognitive abstraction [54]. However, one should keep in mind that recorded foci as well as reifications are *information items*. They are part of a technical language which is intended to *guide* cognitive constructions for practical reference purposes. Even though this language is grounded in cognitive operations, reifications do not imply formally corresponding cognitive counterparts[26]. Furthermore, in order to account for non-extensional constructs, it is possible to allow for non-extensional reifications, as will be discussed in the next section.

Extensional reifications correspond to Bentham's 'aggregation fictions' and Vaihinger's 'summatory fictions' ([197, Chapter 8][27].

Reification of classes. I introduce *reified classes* as first-order entities with an element-of relation $\in$ in the domain of discourse D. F is defined as a predicate denoting urelements (foci of attention). Classes are extensional, so their elements give rise to an identity criterion for classes. The relation $\in$ therefore individuates classes:

Definition 2 *Class definition:*
$Class(x) \leftrightarrow \exists y.y \in x$

Axiom 1 *Classes are extensional:*
$(Class(x) \wedge Class(y) \wedge (\forall z.\,(z \in x \leftrightarrow z \in y))) \rightarrow x = y$

Axiom 2 *A consistent finite class type hierarchy of order $n \geq 1$:*
$C^0(x) \leftrightarrow F(x)$
$\bigwedge_{i=1}^{n} C^i(x) \leftrightarrow Class(x) \wedge (\forall y.y \in x \rightarrow C^{i-1}(y))$
$Class(x) \rightarrow \bigvee_{i=1}^{n} C^i(x)$

[26]Cognitive categories, in contrast, seem rather constructed based on a Gestalt-like auto-associative mechanism or a *simulation*, as described in [64].

[27]In the appendix of the first part in [197], Vaihinger distinguishes analysis ('Zerlegung') from summation ('Zusammenfassung'). The first one can be conceived in terms of Carnap's 'Quasianalse', i.e., distinguishing concrete qualities by making different kinds of classifications [25, §71]. I follow Carnap in arguing that there is really only one method here.

Axiom 3 *F denotes a nonempty set of urelements:*
$(F(x) \rightarrow \neg\exists z.z \in x) \wedge \exists y.F(y)$

Note that since classes are entities different from their elements due to Axioms 2 and 3, Russell's paradox cannot occur. In particular, I only allow for n mutually exclusive layers of classes. It can, e.g., be proved that a class cannot range among its elements. In the simplest case $n = 1$, i.e., there may be only classes of foci, denoted by predicate C^1. Note also that Axiom 3 furthermore leaves open the possibility of logical reifications which are not classes.

The *act of reification of classes*, as discussed in Section 4.1.4, is expressed by an *existential quantification* of an abstract class entity. Observation predicates can be used to define the element-of relation between a class and its elements. We assume there is a finite list $Cl =< \Psi_1, ..., \Psi_n >$, $i \in \{1, ..., n\}$, of unary predicates, for which we have decided to reify their classes. We add for each predicate in Cl one existential axiom of the following form, where Ψ_i needs to be substituted with the ith predicate, respectively:

Axiom Schema 1 *Existence of Ψ classes:*
$\bigwedge_{i=1}^{n} \exists e.\forall x.\Psi_i(x) \leftrightarrow x \in e$

Axiom Schema 2 *There are only Ψ classes:*
$\bigvee_{i=1}^{n} \forall x.x \in e \leftrightarrow \Psi_i(x)$

We furthermore require with Axiom Schema 2 that this is the only way to introduce classes in D. It will be useful to introduce some finite set and mereological operators[28] for these classes (where P stands for "part of" and PP for "proper part of"):

Definition 3 *Class intersection:*
$a \cap b = c \leftrightarrow Class(c) \wedge (\forall x.x \in c \leftrightarrow (x \in a \wedge x \in b))$

Definition 4 *Class union:*
$a \cup b = c \leftrightarrow Class(c) \wedge (\forall x.x \in c \leftrightarrow (x \in a \vee x \in b))$

Definition 5 *Class mereology:*
$P(a, b) \leftrightarrow Class(a) \wedge \forall x.x \in a \rightarrow x \in b$
$PP(a, b) \leftrightarrow P(a, b) \wedge \neg P(b, a)$

A class may also be defined by enumeration of its elements:

Definition 6 *Element enumeration:*
$\{x_1, ..., x_n\} = z \leftrightarrow [\bigwedge_{i=1}^{n} x_i \in z] \wedge \forall x^*.x^* \in z \rightarrow [\bigvee_{i=1}^{n} x_i = x^*]$

[28]Note that I do not consider these as a substitute for mereology in general.

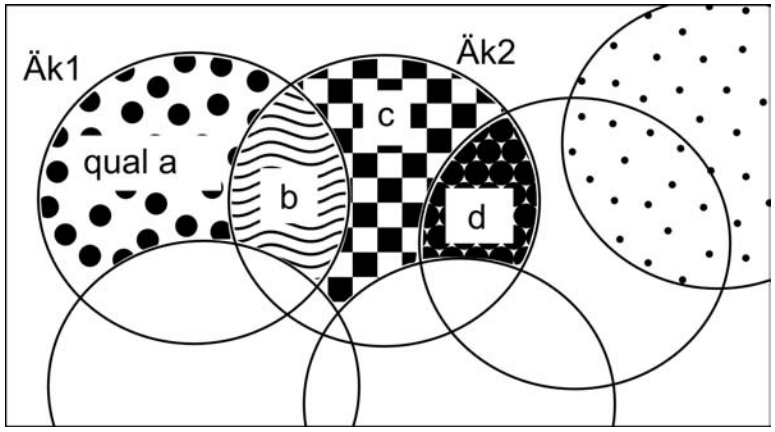


Figure 4.3: Carnap's 'Ähnlichkeitskreise' (Äk, unified wholes) and quality classes (quale wholes) illustrated in a 2-dim space, where a certain maximal plane distance is taken as the similarity relation R [25].

Unified wholes. There is a certain kind of class which is central to the reification of qualities, bodies and other ontological entities. *Unified wholes* are based on *binary predications* and *attentional scope*, and they are *defined in a recursive manner*. In the following axiom schemas, let Φ denote a unary predicate wildcard restricting the attentional scope of the observer on D. Let R be any *binary predicate* on D as introduced in the subsequent chapters, corresponding maybe to some predication on F.

The following axiom schema describes a predicate $Whole_R^\Phi$ which applies to those reified classes that form a single whole w.r.t. Φ and R. All elements of such a whole must be mutually connected by R (this is called *unity*), and the class must be maximal in the sense that every entity R-related to all of the whole's elements is included (compare Figure 4.3):

Definition 7 *Unification:*
$Unity_R^\Phi(e) \leftrightarrow Class(e) \wedge \forall x, y. (x \in e \wedge y \in e) \rightarrow (R(x, y) \wedge \Phi(x) \wedge \Phi(y))$

Definition 8 *Maximality:*
$Whole_R^\Phi(e) \leftrightarrow Unity_R^\Phi(e) \wedge$
$(\forall x. (\Phi(x) \wedge (\forall z. z \in e \rightarrow xRz \vee zRx)) \rightarrow x \in e)$

A $Whole_R^\Phi$ is therefore equivalent to Carnap's 'Ähnlichkeitskreis' [25, §70-71], except that I do not require reflexivity or symmetry of the predicate R, and additionally assume an attentional scope Φ. I will explain the concept of a whole now with respect to different characteristics of R and Φ. These correspond to different actions in terms of the attentional apparatus of Section 3.4.

Perspective and partially covering wholes. An *attentional scope* Φ on wholes is a way of expressing the observer's perspective on a domain and its Gestalts

(compare Sections 3.3.3.2 and 3.4). Wholes without attentional scope allow to construct boundaries simply by ceasing predication, i.e. by the fact that it was not possible to predicate something beyond them. The attentional scope can additionally restrict the perspective on these boundaries, so as to allow the construction of mass nouns (compare Section 3.3.3.2).

Lacking *reflexivity* of R allows predications to cover only parts of the domain[29]. This is crucial, since observation predicates have specific subdomains of F, and predications are usually not performed on the whole subdomain.

Reachability wholes. If R is *not symmetric*, it does not express similarity, but rather *reachability* like in some directed graph. Even though this was not envisioned by Carnap, the concept of such a *reachability whole* can be very useful. It expresses some configuration which is self-reachable from each of its elements but cannot reach and is not reachable from outside. An example of such a whole is a complete road network, as introduced in Chapter 7. In this case, the underlying predication is one of following a *simulated movement* with attention.

Similarity, equivalence and quale wholes. If R is *symmetric*, it can express some *perceptual similarity* in Carnap's original sense. In terms of the attentional apparatus, this means that the observer remembers whether some *Gestalt is present at two focused spots*, without implying any direction. We have to distinguish two subcases, which differ with respect to the possible interpretations of R and the corresponding wholes:

In case R is also *transitive*, similarity corresponds to *equivalence*: $R(x, y)$ then means that x and y "belong to the same thing". More correctly, attention is focused in both cases on one and the same Gestalt with *crisp boundaries*, so that all pairs of foci on this Gestalt are equally well part of it. A whole then corresponds to an R-equivalence class and is called *equivalence whole*. An example for such a whole could be the dalmatian dog in Figure 3.2. In general, such a body Gestalt has a crisp boundary and does not overlap with another one. And neither does the underlying predication expressed in R. If we follow several bodies with attention, our gaze jumps from one point on their body surface to the next one, but it never happens that we are unsure whether a part belongs to the same body or not. It is also clear that all parts belong to the same body to an equal degree. In this case, Definition Schema 8 can be simplified to:

Definition 9 *Maximality (in case R is an equivalence):*
$Whole_R^{\Phi}(e) \leftrightarrow Unity_R^{\Phi}(e) \wedge (\forall x, y.(\Phi(x) \wedge R(x, y) \wedge y \in e) \rightarrow x \in e))$

In case R is *gradual, i.e., non-transitive but symmetric*, it can express similarity as a matter of degree, and therefore wholes can overlap. The underlying

[29]Remember that in contrast to Carnap's single similarity relation, we can have many of them covering different parts of F.

Gestalt in this case has *fuzzy boundaries* and *can overlap with other ones of its kind*. For example, a certain color percept, say, "orange", belongs to some focal color, say "red", only up to a certain degree. Another percept, "yellow", is similar to this percept but may not be similar to the focal color anymore. Each whole in this case corresponds to some overlapping *similarity neighborhood* of colors. Therefore, it is often useful to construct the *maximal whole parts that are not dissected by other wholes*. I call these parts $QualeWhole_R^\Phi$, because they correspond to maximal classes of percepts indistinguishable by the respective predication. One example for such a quale whole is an atomic color value (compare Figure 4.3). Quale wholes correspond to Carnap's 'Qualitätsklassen' [25, §112].

Definition 10 *Quale wholes are maximal non-dissected whole parts:*
$QualeWole_R^\Phi(q) \leftrightarrow \exists e, z.(z \in q) \wedge Whole_R^\Phi(e) \wedge (z \in e) \wedge$
$\forall y.(y \in q \leftrightarrow \forall u.(Whole_R^\Phi(u) \rightarrow (z \in u \leftrightarrow y \in u)))$

Note that quale wholes *are* wholes just in case R is transitive. In this case, wholes cannot overlap. Therefore the elements of the quale whole must be elements of only one whole, namely itself.

Identification of wholes via attention. Since quale wholes as well as equivalence classes are mutually exclusive, they can be identified by any one of their instances, e.g., by pointing towards an ISR spot in attentional focus. This can be done with the following defined function $[x]_R^\Phi$, which maps entities in attentional focus to reified wholes whose existence were asserted by one of the existential axioms above. In the simplest case, if R is symmetric and transitive (a partial equivalence operation), this function simply returns the unique R-Whole in the Φ-subdomain that contains x, given that the whole's existence has explicitly been asserted by a an existential axiom and x is element of it. In case R is not transitive and wholes can overlap, it produces the unique reified quale whole. Otherwise, it produces a special entity denoted by the constant *error*[30]. This reflects the fact that if someone tries to observe colors in darkness, this person will not be able to specify them. Similarly, someone might not recognize the dalmatian dog as an entity. In the first case, the predication underlying the whole is missing, in the second case, the dog was not introduced as a single entity.

Definition 11 *Whole identification:*
$[x]_R^\Phi \equiv Ic.(x \in c \wedge QualeWole_R^\Phi(c))$

Other kinds of 'definitions in use'. Wholes are just one example of a 'definition in use' according to Carnap, i.e., an extensional reification of a class[31]. I will distinguish two main subcategories of such definitions.

[30] This functionality is provided by the iota operator I.
[31] The German term is 'Gebrachsdefinition' [24].

Explicit class definitions. These patterns correspond to an *explicit (non-recursive) definition*, i.e., one in which the definiendum does not occur in the definiens. For example, Definition 6 is explicit. But also, take for example the construction of a triangle. Such a triangle is a class of three linearly independent foci of attention. It is definable by *observed collinearity* among foci: The third focus must not be on a line with the other two. But its definition is not recursive.

Recursive class definitions. All definitions involving *maximality* or *minimality* constraints fall under this heading, and so does the definition of a whole[32]. As we will see in Section 7.3, important examples for minimality patterns in road network databases are junctions.

4.2.2. Non-extensional reifications

Non-extensional reifications can be simply defined as follows:

Definition 12 *Non-extensional reification:*
$Rfc_{ne}(x) \leftrightarrow Rfc(x) \wedge \neg Class(x)$

An important kind of non-extensional reification, which is less strictly dependent on other entities than classes, is *reification by analogy*. These reified entities exhibit some structural similarity with already known classes, but they are not classes themselves. Formal approaches to construct analogies could be *morphisms*, i.e., mappings from class entities to analogical entities that preserve some structural property. Analogies can further be differentiated according to whether *constructed functions connect them to known entities*. If yes, then those connected analogical entities can have *identity criteria* in terms of known classes.

Reference analogies are used in order to imagine *observable but yet unobserved* things, such as times and positions in *temporal and spatial reference systems*. A similar mechanism can account for *object persistence over time* when the object is not observed. In all these cases, we *need to replicate* reifications, i.e., observed classes, but without any actual predications. Furthermore, we need to preserve structural relations to neighboring entities. Therefore, reference analogies have some *definite description in terms of known classes* in order to identify them.

For example, if we construct a temporal reference system like a *calendar*, we need to reference events that are beyond our immediate experience, e.g., dates before our own birth (compare Figure 4.4). The whole system is anchored in experienced events, e.g. daytimes, but extends beyond them by repeating unobserved daytimes in an analogical inductive fashion and by connecting them with an imagined time order[33]. Similarly, we have to construct unobserved locations in spatial reference systems, and the unobserved persistence of objects.

[32]This means that the underlying Ψ-predicate is recursive. In definition 8, e.g., the predicate corresponding to the instances of a whole needs to reappear in the definiens instead of $x \in e$.

[33]Note that this time order cannot be an observation predicate or any predicate defined thereupon. It has to be constructed on top of known experienced entities and is a pure fiction itself,

One important aspect is that the reified analogical entities, since they are associated with the same perceptual operations, can *potentially be observed*, for example in the future. Another aspect is that the reified entities can be *identified* based on known entities and a constructed function, which is an *illegitimate extension* [197] of an observed relation. For example, the "day before yesterday" could be identified by the reader based on a constructed time order relation if he or she knew at which day I wrote this text.

Similarly, we may repeat entities in a *series*, such that they are identifiable based on known classes, but without associating any perceptual operation. This frees us from the idea of observability. Examples are recursive definitions of natural numbers.

It remains a research question how such illegitimate extensions can be formalized and how they can be updated if new experiential evidence is acquired[34]. Following my suggestion in Section 4.1.2, which was based on Lorenzen [100], one possibility is to conceive them as *abstractions from a recursive rule representation of extensional reifications, which were in turn abstracted from perceptual operations.*

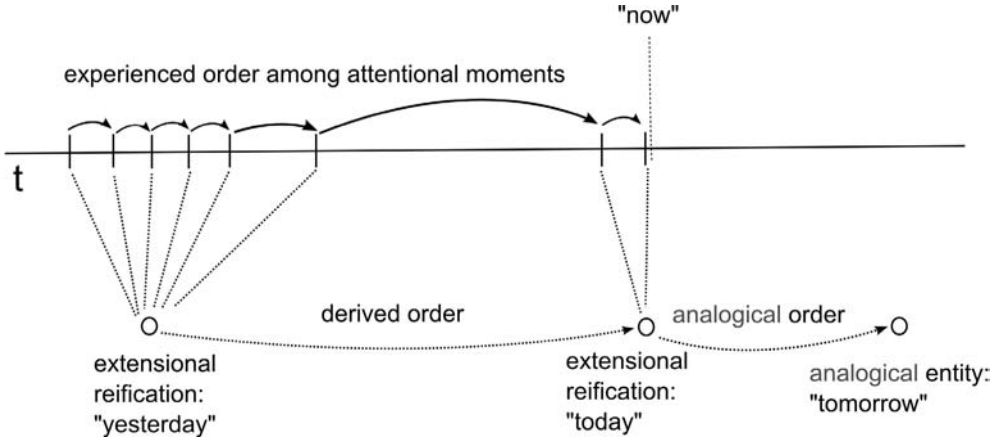


Figure 4.4: Illustration of reference analogies used to construct a personal calendar.

Another less constrained but important kind of analogy are (linguistic) *metaphors* and image schemas, as analyzed by Johnson [74] and Lakoff [90][35]. I suggest to conceive of metaphors simply as morphisms inside of D. Such morphisms allow

in the sense of an illegitimate extension ('unberechtigte Übertragung') proposed by Vaihinger [197, Chapter 7].

[34]For example, when a future day is experienced and the old posited one needs to be replaced in memory. It may appear that this mechanism is similar to a Gestalt mechanism, and thus similarly difficult to analyze.

[35]They were identified as useful fictions already by Vaihinger, disguised in his 'symbolic fictions' [197, Chapter 4]. I guess that other types of Vaihinger's fictions could be subsumed here, e.g., 'die Methode der unberechtigten Übertragung'. This also corresponds to Kant's idea that all human categories can only be based on cognition by analogy [197, Chapter 4].

to structure abstract entities in an analogous fashion, but do not allow to identify them.

I agree with Lakoff that category expansions are arbitrary to a large extent, and therefore graded categories do not have necessary and sufficient conditions. As I conceive it, this may just be a consequence of metaphorical extensions being fictions, and fictions being arbitrary logical reifications involving choice.

Further types of logical reification await discovery. One type may be called *variegated reification* because it purposefully diverts from known structures. For example, deliberate simplifications of reality [197, Chapter 2], as in the case of scientific models. Or to recombine known entities to derive a greater variety of things than given in experience. For example, n-dimensional space is a generalization of experiential 3-dimensional space [197, Chapter 11].

4.3. Reflective abstraction using logical reification

In summary, I have argued in this chapter that a theory of information grounding needs to be about the different ways of constructing abstract entities in thought, and that abstract entities may be conceived as logical reifications. Following constructivists like Piaget and Glasersfeld, but also Lorenzen, abstraction means to reflect on action representations. I have identified 3 possibilities to represent actions and operations, namely through their structured outputs (e.g. memorized relations on foci of attention), through single entities standing for operational schemes, or recursive rules. I have then discussed language tools that can be used to guide abstractions. I have argued with Quine that the tool of existential quantification together with observation predicates, established inter-subjectively in terms of observation sentences by our attentional apparatus, is useful for this purpose. In particular, it can be used to individuate abstract entities based on observation predicates. I have then suggested some kinds of logical reifications and expressed them in FOL. I have introduced extensional reifications as explicit or recursive classes, and distinguished reachability-, similarity-, equivalence-, and quale-wholes as examples for recursive classes. Lastly, I informally discussed some (non-extensional) analogical reifications, namely reference analogies and metaphors, and hinted at formal approaches in terms of recursive rule abstraction and morphisms.

Chapter 5

Perceptual sources for data grounding

The certainty of ideas is not the foundation of the certainty of perception
but is, rather, based on it -
in that it is perceptual experience which gives us the passage
from one moment to the next and thus realizes the unity of time.
In this sense all consciousness is perceptual,
even the consciousness of ourselves.

— M. Merleau-Ponty [113]

In this chapter, I will suggest formal primitives of a grounded first-order theory based on cognition and perception research. These primitives are suggestions for *observation predicates* to be established in the way described in Section 4.1.4. I argue for these predicates based on certain shared perceptual operations. Remember that observation predicates are reference tools of a technical observation language. They are established among observers in order to refer to the output of shared perceptual operations (Section 4.1.4). This output is stored in the observer's memory. Perceptual operations, in turn, become manifested in quasi-"universal" capabilities for perceptual predication and referencing in the perceived space around the body. Predications are performed by applying Gestalt mechanisms to the ISR domain (compare p. 34) and focusing attention on those Gestalts (compare Section 3.4).

It is essential for a grounding theory that there are observations which can be reproduced *independently of who performs them and how they are interpreted.* I argued in Section 3.2 with Pylyshyn that human reason would be captured in a closed cycle otherwise, and that therefore such operations need to exist. I suggested that humans can easily understand and reproduce observations made by others, because they can understand intentions and join their attention in a scene (Section 3.3.3.1). If someone tells you that Main Street is closed due to construction works, you can easily understand and reproduce what was observed.

This simple fact was recognized by Merleau-Ponty, who claimed the *primacy of perception* in cognition [113]. This idea contrasts with the standard paradigm of

cognitive science, which could be called *the primacy of sensation*. This latter view considers perception an intellectual product of interpretation of atomic sensations, and therefore falls prey to the subjectivity of thought. Instead, Merleau-Ponty argued that pure sensations are fictions of the scientific mind, whereas it is Gestalt perception that builds the reliable ground for all conscious thought [113].

This means we need to look for perceptual predications which can be performed largely independently from varying mental concepts or beliefs. The experiences have to emerge spontaneously, i.e., as a result of drawing the attention to externally triggered mechanisms, or of performing some other simple attentional operation. The results of such operations therefore have a bottom-up priority over conceptual reasoning (compare Section 3.2). They still may be influenced by control of attention with certain concepts in mind. Also, the operations themselves may be a result of cultural learning (as argued in Section 3.3.3.1). Nevertheless, they must be based on some pre-conceptual mechanism, as for example a Gestalt mechanism. This makes them an anchorpoint for establishing language references.

I will first discuss perceptual operations and introduce observation predicates in Section 5.1. Quantifiers in all axioms are assumed to range over the whole universe of interpretation D if not explicitly restricted to attentional moments F[1]. The suggestions made indicate *perceptual sources* for a meta-theory of grounding, i.e., for a theory that allows to build grounded theories, and thus semantic reference systems, in an information community. Some of them were already proposed by myself in [159]. I conclude in Section 5.2 by discussing metalogical properties and summarizing the scope of the proposed method of grounding.

5.1. Sources for perceptual predications

In this section, I review empirical evidence for the existence of predications that allow to relate entities in attentional focus. These can be used to construct and individuate qualities, objects, actions and media in the domain of intermediary spatial representation (ISR). *Observation predicates are symbolized by uppercase letters*. I will describe them by defendable axioms and axiom schemas in this and the following chapters.

5.1.1. Attentional moments

I agree with von Glaserfeld[2], that there has to be some "pulsing" mechanism that produces *discrete* mental entities on the very lowest level of conscious perception[3].

[1] For convenience, I will indicate such a restriction by the quantifier $\forall x \in F$.

[2] v. Glasersfeld developed a 'pulse' model for the mental construction of unities, pluralities and number (see [195, Chapter 9] and [194]).

[3] Although the question of whether conscious perception is discrete or not is in principle still open, there is much psychophysical evidence for its discreteness [198].

I assume the identity of a moment in which a human being focuses its attention on a certain signal from the space around the body. This signal is a Gestalt precomputed on the ISR domain. The ISR domain and its Gestalts are pre-conceptually synthesized from visual, tactile, proprioceptive and other signal inputs, without the observer being necessarily aware of them (Section 3.4). They enter consciousness only via *attentional moments*. The domain of attentional moments was already introduced in Section 4.2 and called F. It is considered as a kind of root for other domains of consciousness, as it is the only mental domain which can be directly coordinated across observers by the mechanism of joint attention (Section 3.3.3.1). It is also considered to be finite and therefore discrete, because human memory is bounded.

Observation predicates. One *focus of attention* can be distinguished from another one because they come at different discrete attentional pulses. These pulses can be considered a temporal Gestalt mechanism. Perceiving time in its simplest form therefore means to perceive the *temporal order* $\leq_T$ of foci of attention in F. Because I assume that attentional moments are indivisible for an observer, this order relation accounts for *true identity* (denoted by $=$) in F[4]. The order is total in the sense that every two moments in F are comparable.

Axiom 4 *Temporal order with identity:*
$(F(x) \wedge F(y)) \leftrightarrow (x \leq_T y \vee y \leq_T x)$ totality
$(x \leq_T y \wedge y \leq_T x) \rightarrow x = y$ antisymmetry
$(x \leq_T y \wedge y \leq_T z) \rightarrow x \leq_T z$ transitivity

In a finite poset an element with a successor always has an *immediate successor*. One can define this immediate successor relation $\prec_T$ by first defining a strict order.

Definition 13 *strict order:*
$x <_T y \leftrightarrow x \leq_T y \wedge \neg y \leq_T x$

Definition 14 *Immediate successor:*
$x \prec_T y \leftrightarrow x <_T y \wedge \neg(\exists z.x <_T z \wedge z <_T y)$

Since the domain of foci denoted by F is assumed to be finite, every focus with a successor/predecessor must have an immediate successor/predecessor. I spare the axiomatization of this here. In the following, I may write $Whole_R^F$ in order to denote R-wholes[5] that are unrestricted on F.

Note that the *conception of time* exceeds this raw mechanism of pulsing attention in terms of a temporal metric. For this reason, time measurement needs

[4]This is a pragmatic assumption. It may be necessary to allow for more than one synchronous focus of attention at a time. Pylyshyn [135] argues for four of them. This can be done by assuming an identity relation for attentional moments distinct from temporal equality.

[5]See p. 80 for the introduction of a whole.

to rely on some other temporal reference mechanism that assures uniformity of intervals, as for example the ticking of a clock.

The pulsing attention does not have to be focused on another signal, e.g., a Gestalt. If it is, it produces a constant flow of conscious experience. The mental operations discussed in Section 4.2 are then available to construct higher level entities from this material flow of consciousness. As argued in Section 3.4, conscious experience simply means that human observers detect and store the presence of a Gestalt structure at one or several *foci of attention*.

5.1.2. Perceiving the meaningful environment

The ecological psychologist Gibson [50] suggested an informal ontology of elements of the environment that are accessible to human perception and action, called the *meaningful environment*. Gibson's proposal turns out to be very useful as a source for grounding, even if one does not follow his epistemological theory of 'direct perception'. It serves as a guiding principle to investigate empirical evidence for perceptual operations for all of its elements. The three top-level categories of *meaningful things* [50, page 33] in this environment are *substances*, *media* and *surfaces*.

A *medium* affords moving through it as well as seeing, smelling and breathing, and it bears the perceivable vertical axis of gravity (for vertical orientation). According to Gibson, the medium for terrestrial animals is the air. Gibson thought of a medium as something established in terms of *affordances*, i.e. action potentials in the environment. For example, he distinguished liquid media (water) and gaseous ones (air) by what actions they afford to the animal [50, Chapter 2]. I have suggested in [158] that there may be different kinds of media according to what kind of action they offer to a human being. In this thesis, I restrict my understanding of a medium to those based on locomotion and action affordances. This view will be explained in Subsection 5.1.7.

Surfaces are the boundaries of all things humans can distinguish by visual perception. This means they are opaque to a certain extent and they bound an *illuminated medium*, i.e., a medium for seeing, such as air. Surfaces have *surface qualities*, for example a *texture* (including color), and are often resistant to pressure.

Substances are things in the environment that are impenetrable to motion (i.e., are solid) and illumination (i.e., are opaque). Detachable substances are called *objects*, which have further properties, e.g. a shape and a weight. Moreover, substances enable actions: they support movements (as ground), they enclose something (as hollow objects), or they allow to be thrown (as detached objects).

The identification of surfaces in terms of Gestalt mechanisms is discussed in Section 5.1.3. A possible construction of substances and objects/bodies is proposed in Section 6.3.2.

One of Gibson's central insights was that the elements of the meaningful

environment are inter-subjectively available to human observers in their domain of experience. However, if one does not assume that observers have direct access to external reality[6], this can only mean that they have analogous capabilities for *identifying and distinguishing* these things. I suggest in the following that meaningful things should rather be viewed as results of mental constructions based on pre-conceptually available Gestalt mechanisms. For example, bodies are perceived based on a surface Gestalt mechanism. Complex qualities of bodies can be constructed by performing perceptual operations on their surface layout, e.g. by constructing their lengths or depths [158]. Movements and other events can be individuated by following these bodies with attention [159]. Media can be individuated based on perceived affordances [158, 156]. For example, the affordance of locomotion identifies a medium that allows you to move through. This can be just the free space of your office, when the door is closed, or extend several kilometers throughout the landscape when you are hiking outside.

5.1.3. Identifying visual surfaces

Mainstream philosophy of the mind says that *objects* (in general: particulars) are constructed from more primitive 'point-like' percepts of qualities using conceptual reasoning and the application of knowledge[7]. As argued in Section 3.2, such a view gives rise to the philosophical puzzles that concepts imply other concepts in infinite regress, and of how to explain that humans can share knowledge although equivalent concepts cannot universally be expected.

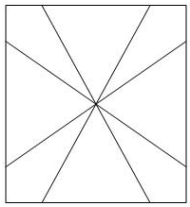


Figure 5.1: Slender x or thick cross? Cf. [80].

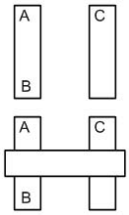


Figure 5.2: 'Same object advantage' in attentional spread, cf. [161].

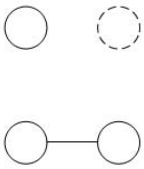


Figure 5.3: The effect of 'simultanagnosia', see text for details.

A remarkable exception to this view was taken by some Gestalt Psychologists, for example Wolfgang Köhler [80]. He realized that the perception of objects in its most general sense is not an epistemic issue, but the result of an unconscious, dynamic self-distributed process, which is hard-wired into the perceptual system, and which spontaneously cuts out units from the signal input which correspond

[6]This view, called 'direct perception', is often held among ecological psychologists [168].
[7]See for example Strawson's influential idea of *feature-placing* [179].

to identifiable surfaces of the environment. One classical example is the *figure-ground distinction* depicted in Figure 5.1. The picture can either be perceived as showing a slender x or a thick cross – but not both at the same time, even though the information in the figure does not preclude this. If we decide to consider a spot to be part of one object (the 'figure'), then the ground is automatically identified.

This mechanism is at the same time an autonomous apparatus for individuation of *visible surfaces* in the environment, and of other things based on them. Note that Gibson's proposal was explicitly based on the insights of Gestalt psychologists. He viewed surface perception in terms of an ecological constraint of the visual system [50] which is basic for many other concepts. For example, he conceived an object in terms of a detachable topologically closed surface.

Empirical evidence. There is recent empirical evidence for such a mechanism coming from studies of object-based focal attention (see Scholl's survey [161]). Same object advantages for example were demonstrated in target detection games. If attention is drawn to one end A of the bar-like object depicted in Figure 5.2, then *same object targets*, e.g. a sudden luminance decrement at B, can be detected faster than targets at C with exactly the same distance. The effect is stable even if the bars are partially occluded by another bar. There is an automatic spread of attention along the outlines of what humans perceive as an object, which takes into account its partial occlusion, its topological connectedness, and seems to be symmetric and transitive. Further evidence for this mechanism is provided by the *Balint* syndrome, which is an object-based perceptual disorder that restricts attention to only one object at a time (*Simultanagnosia*). While Balint patients are typically unable to see two separate discs simultaneously on a screen (see upper half of Figure 5.3), they are perfectly able to see a single dumbbell (see lower half of Figure 5.3), which suddenly appears when the two discs are connected by a line [161]. A similar mechanism seems to stand behind Pylyshyn's [135] FINSTs. As I mentioned already (Section 3.2), FINSTs enable humans *to track* up to four such moving objects in a complex dynamic scene without any decrease in performance. This means that surface individuation is also responsible for keeping object identity in a scene and under perspective change[8].

Observation predicates. What is sometimes called the *individuation of a body in time and space* in the philosophical literature, can be conceived as an *equivalence whole* (compare Section 4.2) connected by a *partial equivalence relation* on F. This relation is called *surface-connected* (SC). It practically solves the philosophical problem of *object identity in time* by detecting whether two foci of

[8] I furthermore assume that this mechanism decides whether a strongly connected body stays the same if it loses parts, e.g. breaks. The mechanism therefore cannot be restricted to mere topological properties, compare [18].

attention focus on the same surface [9]. In this way, humans are spontaneously able to 'cut out' those parts or regions in the flow of conscious experience which correspond to the same body or the same surface *irrespective of its other perceived qualities*. The underlying operation cuts out just those parts of the visual field which correspond to Gibson's [50] 'solid angles in the ambient optical array', that is the surface patches in the perceivable environment. The relation has to be a *partial* equivalence relation (so it is not reflexive), because not everything is a focused part of a body.

Axiom 5 *Surface Connection:*
$xSCy \rightarrow ySCx$ symmetry
$xSCy \wedge ySCz \rightarrow xSCz$ transitivity

The interaction of this predicate with others and the construction of bodies based on it are described in Section 6.3.2.

5.1.4. Identifying point-like and other features

Surface perception plays a central role for many other kinds of perceptual operations that can be performed. In this spirit, Gibson [50] granted surfaces a central position in his ontology. I argue in this subsection for a very important (but often overlooked) kind of surface-based perceptual predication. It allows to identify *features in the environment with respect to some host surface*.

Observers identify prominent parts of their environment, such as relative parts of bodies, openings, or the free space in front of them, with respect to some already identified reference surface. These things are called *features* in the DOLCE ontology [109]. They have their own criterion of identity, but existentially depend on an identifiable object, which is their 'host'. While a feature needs a host, it does not need to be part of it. Perceivable features of a cup, for example, are its handle but also its opening. The opening of a cup would not exist without it, but is not a part of the cup. A feature of a building is the opening of its entrance. One may call openings *open features*, because they are part of some medium, and handles *closed features*, because they are part of some substance. Further examples for features are the corner of a table or the peak of some mountain. I propose to call these latter examples *point-like features*, because they are based on concentric sphere Gestalts that correspond to the mathematical fiction of a point. As I will argue in the next subsection, these kinds of features play a particularly important role in establishing spatial reference systems. There are also linear features, like edges of a table (compare Gibson's [50] treatment of visual edges), as well as features of other shapes.

[9] Note that this mechanism cannot account for identity across long intervals of spatial absence. General criteria of object identity are much more complex and may often involve human reasoning.

Empirical evidence. Features are an important class of perceivable entities on their own, even though they depend on host surfaces. I assume that feature predications are based on a sort of Gestalt mechanism which is triggered by a certain configuration of perceivable surfaces.

Studies in Gestalt psychology bear evidence for some sort of visual "hidden structure" that may account for this phenomenon. Rudolf Arnheim [3] studied the visual perception of balance, shape and form. He noticed that the perception of balance of black dots drawn into a square (Figure 5.4) depends on how they are placed relative to the hidden field of visual tension shown in Figure 5.5, which emerges relative to the square. Note that this field is not part of the square drawing. It rather depicts how black dots in the square are "dragged" towards its centers by a field of visual force. Arnheim assumes that this Gestalt mechanism accounts for the apparent human ability to detect whether the black dot is slightly off-center, without consciously comparing directions and lengths.

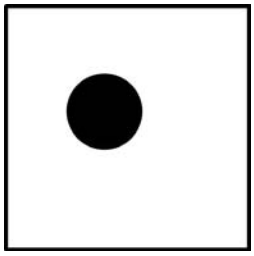


Figure 5.4: A dot placed off-center into a square, cf. [3].

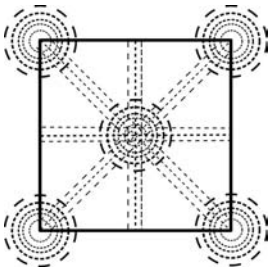


Figure 5.5: The hidden field of visual force exerted on a dot placed into a square, cf. [3].

Attentional studies show that relative parts of objects as well as locations in the background environment can be easily identified [161] relative to reference surfaces. There is also recent evidence for a neurological mechanism underlying the intuitive sense of a location [22]. Burgess and others studied neurons in mammals, e.g. rats, that identify relative allocentric places and called them *place cells*. These cells fire in response to other cells (called *boundary vector cells*), that detect surfaces at a certain allocentric direction and distance (see Figure 5.6). Allocentric means that the firing of all these cells is independent of an egocentric reference frame, but depends on external landmark objects and surfaces [22]. If a rat comes across a place defined relative to the walls of a box, the cell will fire regardless of the direction of approach. There are even place cells configured in a grid-like manner [22]. Therefore, point-like features of this kind may be called *proto-locations*. They can be considered preliminary elements of spatial reference frames.

The space of kinesthetic coordination of our body relies on lots of similar allocentric and egocentric mechanisms. These allow pre-conceptual localization

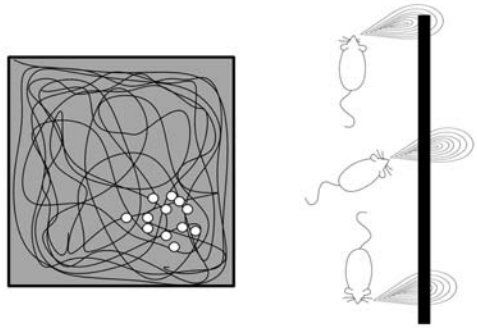


Figure 5.6: Place cell firing (white dots) of a rat tracked in a box (left). Principle of boundary vector cells (right), cf. [22]. See text for details.

relative to diverse body parts. Rizzolatti [145, Chapter 3] and others have described bimodal neurons in the premotor cortex of apes that react to visual as well as tactile percepts. These neurons fire if an object touches some field of skin on face, neck, or extremities, but *also if it is observed moving into some relative geometric corridor in front of it*, regardless of the direction of view. Rizzolatti supposes that such autonomous local reference mechanisms liberate the observer from computing complicated coordinate transformations [145]. They also provide the physiological ground for intuitive localization of things in 'peri-personal' space, i.e., the space reachable by the extremities[10].

Observation predicates. For reasons of simplicity, I suggest a single ternary observation predicate $(r)PF(x, y)$ for point-like features[11]. The foci x and y in this relation express that an observer focused on the same point-like feature at two moments of attention. The position r denotes an attentional moment focused on the host of the point-like feature, i.e., on its surface. I assume that a feature in general has only one such reference surface[12]. Therefore, r can be used to identify the host of the feature. The induced binary relation $XF^*(x, y) \leftrightarrow \exists r.(r)XF(x, y)$ is a partial equivalence relation.

Other useful kinds of features are assumed to satisfy the same properties. They can be introduced accordingly by substituting some letter in the position X, such as P, in the following Axiom Schemas:

Axiom Schema 3 *Feature Symmetry:*
$(r)XF(x, y) \rightarrow (r)XF(y, x)$

Predicating a feature is not dependent on the order of foci.

[10]This space seems dynamically extensible by way of tools. This means that neuronal corridors can even readjust to "prosthetic" enlargements [145].

[11]Alternatively, one may want to differentiate different kinds of point-like features.

[12]This may also be an oversimplification.

Axiom Schema 4 *Feature Transitivity:*
$(r)XF(x,y) \wedge (r)XF(y,z) \rightarrow (r)XF(x,z)$

Features of the same type cannot overlap.

Axiom Schema 5 *Host Uniqueness:*
$(r)XF(x,y) \wedge (r')XF(y,z) \rightarrow rSCr'$

A feature of a given type has a unique host.

An equivalence class, i.e., a whole $Whole^{F}_{XF^*}$ with respect to the induced relation $XF^*(x,y)$, denotes one single feature. The relation is not reflexive because not every moment of attention is focused on a feature. The connection of features to their host surface is given by r and SC (i.e., surface connectivity, see p. 92). By the third axiom, it can be proven that features have (extensionally) unique hosts, since hosts have unique surfaces.

5.1.5. Identifying locations and geometric properties

In addition to these low level mechanisms, there are also conscious operations available that we ordinarily think of when measuring the geometry of our environment, such as determining lengths, widths, heights and directions. However, on what kinds of primitive experience are such operations based? The following discussion of *experiential geometry* was published in a similar form in [155].

Geometry is a traditional topic of mathematics. As all traditional mathematics, it evolved from concrete experiences and problems. In arithmetics, the experiential basis of counting soon became extended in order to incorporate infinity in terms of the natural numbers, rationals, and reals. Just as mathematicians needed to withdraw from performing explicit countings in order to close the numbers with respect to arithmetic operations, mathematical geometry closed its experiential domain of measurement by assuming infinities of points.

From the perspective of information grounding, the merit of domain closure makes it difficult to see what the roots of geometry are, but they play an important role in reference systems. Spatial reference systems are established by geodesists in terms of observed directions, angles and lengths. Also, many observed qualities have their roots in observed geometry, such as widths, heights, depths and distances. But the kind of geometry performed by a geodesist is essentially different from mathematical geometry in that it is *constructive* and *finite*[13]. This remains true even if calculations are performed on discrete approximations of real number fields, as in computers.

[13]For similar reasons, Habel [61] has proposed that cognitively adequate temporal reference systems should be *finite* with a so-called *density in intensio*. Even though cognitive temporal intervals should be considered finite, this should account for the fact that every time interval *can potentially be bisected*.

I propose in the following that the identity of point locations, the domain of observed geometry, may be constructed based on length and direction comparisons taken with reference to some anchor frame consisting of point-like features, such as a particular end of a bar. This is so because there is no way of determining an absolute location in space and time[14]. I propose that this reference frame can be any perceivable point-like feature configuration which retains certain intrinsic geometric relations.

I suppose to conceive of experiential geometry in terms of predications as well as constructive operations as introduced in Section 4.2. An observer *is able* to explicitly introduce a finite list of locations in his or her domain of discourse based on comparing foci of attention. The existence of locations depends on the observer's choice of taking perceptual as well as constructive actions, and on the choice of point-like features which can play the role of reference points.

Constructive geometries. Patrick Suppes has pointed out that *closure conditions* in modern mathematics, even though in themselves useful tools, can be obstacles in search of computationally feasible and practical solutions to applied problems [181]. I would add that they also tend to hide the experiential roots of mathematics. Similar to Greek mathematics, Suppes' proposal [181] uses constructive finite formalisms in order to deal with applied problems such as how to construct buildings in architecture. In his formalism, geometric figures are explicitly constructed by a finite series of steps from a small point basis. The operations he proposes are *doubling* and *bisecting* of lines, which allow, for example, to construct parallelograms (Figure 5.7).

Note that this approach differs from *finite geometry* in mathematics [102]. One does not check whether a given axiomatization has finite models, such as affine plane figures of finite order. One rather describes *how geometric figures of a Euclidean flavor* can be constructed in finite sequences. This is not feasible in standard formalizations of Euclidean Geometry, since they require infinite models due to closure axioms.

Infinity axioms[15] take the form of *universal-existential* sentences, i.e., $\forall x_1, \ldots . \exists y_1 \ldots . \Phi(x_1, \ldots, y_1, \ldots)$. They allow to express recursions of existence claims, and thus to populate the domain of interpretation infinitely. Such axioms abound in geometry and enforce their models to be infinite. If we take Tarski's elementary axiomatization of Euclidean Geometry [186], then we find four of the

[14]Even if we use a *spatial reference system*, this system is logically anchored in and therefore presupposes the identity of concrete places. Such an anchor place is a necessary part of a *geodetic datum* for a mathematical ellipsoid representing the earth surface.

[15]By this notion, I vaguely refer to the axiomatic causes of infinity in a theory. These may resemble the *axiom of infinity* of ZF-set theory, which enforces a set containing successors for all its elements. I am yet unsure how to make this notion more precise, since universal-existential form is only a necessary criterion. However, I will give examples for infinity axioms in the following.

13 axioms to be of such a form (or translatable into such a form, see [187]). For example, the *axiom schema of continuity (Axiom 13)* requires a boundary point for every two predicates that divide a ray into two halves. This is essentially the idea of a *Dedekind cut*, and thus requires the *continuum* of cardinality $2^{\aleph_0}$.

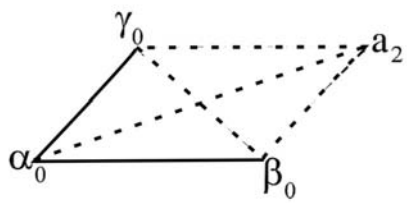


Figure 5.7: Suppes' constructive geometry can be used to construct parallelograms. Bisecting line γ_0, β_0 yields new point a_1 and doubling line α_0, a_1 yields a_2, cf. [181].

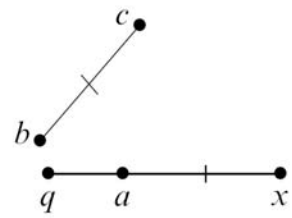


Figure 5.8: Tarski's Axiom of Segment Construction, cf. [187].

But even if we dispense with Axiom 13, it is provable that models still need to be isomorphic to vector spaces over ordered fields [186], and these are infinite on the cardinality level $\aleph_0$. Reasons for this are the three remaining infinity axioms of *Pasch*, *Euclid* and the *Axiom of Segment Construction*. The latter axiom, for example, requires, for any existing pair of points b, c and for any given line (denoted by another pair of points, q, a), the existence of a pair of points a, x on that line which is congruent to b, c (compare Fig. 5.8). This requires infinity by itself: There is now a new pair of points on a line, for example q, x in Fig. 5.8. If we apply the axiom again to this pair and the line pair q, a, it requires a new pair a, x^* congruent to q, x, and so forth. Something equivalent is also enforced by the Axiom of Pasch.

In a constructive geometry, infinity axioms like the axiom of segment construction need to be replaced by *explicit constructions*. These can be expressed in first order logic (FOL) by a finite list of existential quantifications that state the existence of any constructed point. Another possibility is to describe the underlying operations explicitly, not only their results, in the spirit of Piaget's logic [126]. This can be done in terms of a constructive calculus [98], such as those used in intuitionistic logic or algebra, and was demonstrated in [155]. However, in this thesis, I will stick to the level of FOL theories for pragmatic reasons[16].

In this way, it becomes possible to describe finite models that resemble Euclidean geometry in FOL, while allowing to capture the *intention* of an infinite geometry. Points then do not exist unless they were explicitly introduced by constructive operations. In consequence, universal theorems $\forall x_1, ..., x_n.\Psi(x_1, ..., x_n)$,

[16]In order to avoid intuitionistic challenges. These need to be addressed in a separate place, see [155].

provable with the help of closure axioms in Tarski's geometry [187], may be provable only in an *existentially conditioned form*:

$$\forall x_1, ..., x_n. \, [\exists z_1, ..., z_m. \Omega(z_1, ..., z_m)] \rightarrow \Psi(x_1 ... x_n)$$

Here, the condition of the implication requires the existence of some *auxiliary points* $(z_1, ..., z_m)$ in order to prove its consequence[17].

Observation predicates for constructive experiential geometry. Since the operations of doubling or bisecting of lines [181] seem too restricted to capture constructive observed geometry in the space around the body, I will not directly use Suppes' proposal, even though I stick to his approach. Instead, I will develop a constructive modification of Tarski's axiomatization [186], because it is possible to interpret his primitives [187] in an intuitive way.

I have proposed in [158] that humans experience the geometrical and topological structures of their environment by performing and comparing *attentional steps*. An attentional step is the *actual movement of attention* from focus x to y. Humans perceive *length* and *direction* of steps, because they are able to *identify steps of equal length and of equal direction*. And thereby, they are able to observe and measure lengths of arbitrary things in their environment.

I have suggested in [155, 158] that there are (at least) two Gestalt mechanisms available for geometric predication. One is a mechanism for comparing distances between pairs of foci. It can be conceived as the result of using a straight stick or some imagined straight Gestalt and matching its ends with two pairs of foci. For example, physically, we may align a stick with some object and move it around to match with some arbitrary foci of attention. We do exactly this when we use a non-collapsible compass in order to construct a circle. Note that the operation of identifying steps of equal length my be different from the one for performing steps[18]. The observation predicate $xy =_L uz$ (see Fig. 5.9) asserts that foci x and y and foci u and z could be matched in this way[19].

Another Gestalt mechanism allows for perceiving whether three foci of attention lie along a line. $OnL(x, z, y)$ means that a focus z is *on a line* between x and y or co-located with any of them (compare Fig. 5.9). Note that OnL implies collinearity and betweenness[20]. It may be the result of comparing a focus of attention with two others by detecting whether or not it lies on an imagined line through them.

[17]This sort of theorem will be explained and used in Section 6.1 in order to prove metric properties of the proposed formalism.

[18]I assume here that some operation for performing steps generates foci of attention, while some operation of comparing them generates geometric relations between them.

[19]This predicate was called 'congruence' by Taski in [186].

[20]This predicate was called 'betweenness' by Taski in [186].

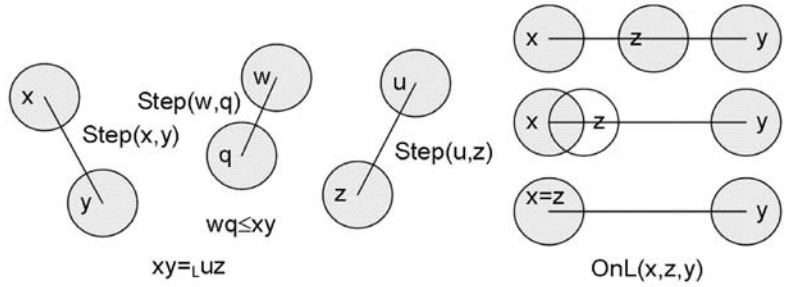


Figure 5.9: Equal length and spatial linearity for steps.

Identification of locations. In distinction to common axiomatizations of point geometry, such as [186], the behavior of these observation predicates needs to be described not on the level of their domain and range, i.e., on the level of foci of attention, but with respect to constructed locations. The "points" of an experiential geometry are locations, not foci of attention. They exist only relative to a *frame of reference*[21] and certain comparison operations, as the ones introduced above. This distinction is ontologically essential, since a given attentional focus can be used to identify different locations with respect to different frames of reference. For example, if one sits in a train and focuses two times on the apex of a table in front, then, at both moments, one is focusing *on the same point* with respect to the table, but on *two different points* with respect to a frame of reference located outside the train and being at rest relative to the landscape.

A reference frame is not only necessary to fix the measurement units and directions of an observed geometry. Together with basic comparison operations discussed above, it actually *establishes a geometry* with its points and its laws in the first place[22]. As A.S. Eddington argued, we must recognize "that all our knowledge of space rests on the behavior of material measuring scales", and not on some pre-experiential absolute space [36].

To illustrate this argument, suppose you are sitting in this train with some measuring tape at your disposal. Focusing on the table in front of you, you can move your attention from one of its ends x to another y and back to the first one x'. You will thereby notice that the length of the table has remained equal, i.e., $xy =_L yx'$, and therefore length comparison with reference to this frame and the tape is seemingly symmetric and suited for Euclidean geometry. But what if you choose a reference frame consisting of the table edges and a tree rushing past the window? If you jump with your attention from this tree to the table edge and back, symmetry of length is not preserved. So the choice of the frame of reference influences formal properties of your geometry. Similarly, if you choose to make

[21]We use the term frame here not in the sense of a formal reference system [81], but in the sense of perceivable point-like features one can refer to.

[22]This idea of relative space was proposed already by Leibniz [95] and Poincaré [128].

length comparisons with a rubber band rather than a tape, symmetry properties may be preserved in the second case, but not in the first one (compare also [36]). So it is the choice of reference frame and comparison operations together that constitute an experiential geometry[23].

Euclidean-like geometry in perceived space can only be established based on choosing a stable reference frame of four *reidentifiable points* for three dimensions. Note that these four points need not only be reidentifiable by a single observer. If the geometry needs to be shared among people, the points also need to be indicated to others. From all I said above, this means they need to be chosen on the basis of shared Gestalts *external* to the geometrical system. I propose therefore that reference points may be based on point-like features, such as one identifiable corner of a perceived table.

Geometrical reference frames and loci of attention. Locations are not primitive notions in our theory. They are defined in a rule-based manner (compare [155]), as illustrated in the following. For convenience of readability, in this paragraph we restrict all free variables to range over the domain of F.

I first define a binary equivalence predicate $x =_{Ref} y$ which expresses whether two foci of attention x and y are *focused on the same location* relative to the reference frame established by four point-like features, indicated by foci s, a_1, a_2, a_3. These features are mutually distinct and configured in a perpendicular manner, as defined below by the predicate *RefFrame*. Note that the predicate *RefFrame* refers to a *particular* reference frame, i.e., a particular configuration of point-like features, which is taken as a standard in this thesis. In spatial practice, there are usually many different frames of arbitrary complexity (and, thus, many different spatial reference systems).

The equivalence classes of $x =_{Ref} y$ are called *loci of attention*. Loci of attention are the points of a geometry built relative to this frame. They are established based on this frame and the observation predicates for comparing lengths and directions. An observer focuses on the same locus with respect to this frame at two opportunities x and y, if and only if both have the same distance to the four point-like features of this frame [155]:

Definition 15 *Identifying loci of attention:*
$x =_{Ref} y \leftrightarrow$
$(\forall s, a_1, a_2, a_3.RefFrame(s, a_1, a_2, a_3) \rightarrow xs =_L ys \wedge \bigwedge_{i=1}^{3} (xa_i =_L ya_i))$

Furthermore, location equivalence $x =_{Ref} y$ means that x and y are bound to have the same distances to every other location in focus[24]. To put it in another way:

[23]But nevertheless, these choices do not yet determine it, as Poincaré argued [128]. Geometry is likewise affected by Quine's empirical indeterminacy [136], in the sense that, given a reference frame and comparison operators, there is more than one way of building a geometry.

[24]This essentially restricts the number of spatial dimensions to 3.

length comparison behaves neutrally with respect to foci on the same location (compare [155])

Axiom 6 *Locus neutrality of* $=_L$*:*
$x =_{Ref} x' \wedge y =_{Ref} y' \rightarrow xy =_L y'x'$

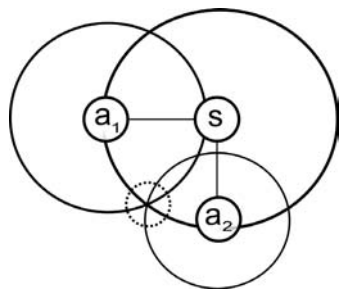


Figure 5.10: Definition of arbitrary point locations in a 2-dim. geometrical reference frame by the unique intersection of 3 circles around reference points. Compare Definition 15.

This definition of location essentially requires that location-equivalent foci lie in the intersection of four spheres that have the point-like features of the reference frame as centers (compare Figure 5.10 for the analogue 2-dimensional case). Every time an observer focuses on a spot in the environment that lies in this intersection, he or she focuses on the same location relative to this frame. That this location is indeed unique and the defined relation is an equivalence relation can be shown based on these definitions, the definition of *RefFrame*, and the Axioms in Section 6.1, compare [155]. In case one of these spheres has a zero radius, the established locus is just coincident with the respective point-like feature. So the features of the frame range among the loci of the established geometry.

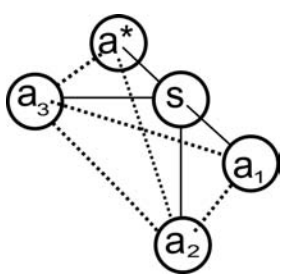


Figure 5.11: A reference frame for a 3-dimensional cartesian coordinate system. Compare Axiom 8.

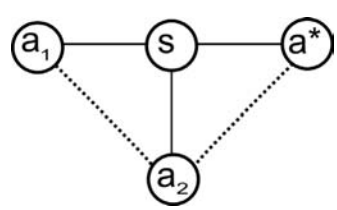


Figure 5.12: The 2-dimensional case of a geometrical reference frame. Orthogonality is enforced by the condition that dotted lines have equal length.

The four point-like features of a reference frame need to retain certain metric and directional relations through time. The predicate *RefFrame* describes such a frame, where s denotes a focus on the origin, and a_i denotes one of three foci on *perpendicular unit vectors* in this reference system (compare Figure 5.11). Focus a^* is an auxiliary focus which is needed to assert orthogonality. Orthogonality is assured by the condition that distances of foci a_i to each other are all congruent, and by the fact that a^* is not on the same point-like feature as s by Axiom 8. That this assumption holds can be inspected for two dimensions in Figure 5.12 and can also be proved[25].

Axiom 7 *Reference frames consist of point-like features:*
$RefFrame(s, a_1, a_2, a_3) \rightarrow$
$(\forall x_s, x_1, x_2, x_3.PF^*(s, x_s) \wedge \bigwedge_{i=1}^{3} PF^*(a_i, x_i) \leftrightarrow RefFrame(x_s, x_1, x_2, x_3))$

Axiom 8 *Configuration of point-like features in reference frames:*
$RefFrame(s, a_1, a_2, a_3) \rightarrow \exists a^*.\neg PF^*(a^*, s) \wedge \bigwedge_{i=1}^{3} \neg PF^*(a_i, s) \wedge$
$OnL(a^*, s, a_1) \wedge \bigwedge_{i=1}^{3} a^* s =_L s a_i \wedge \bigwedge_{1 \leq i < j \leq 3} a_i a_j =_L a_1 a_2$

Axiom 9 *Immovability of reference frame:*
$RefFrame(s, a_1, a_2, a_3) \rightarrow$
$(\forall x_s, x_1, x_2, x_3.RefFrame(x_s, x_1, x_2, x_3) \leftrightarrow x_s s =_L ss \wedge \bigwedge_{i=1}^{3} x_i a_i =_L a_i a_i)$

This last rule says that foci on the same point-like feature of the frame have zero distance from each other. This assures that the reference frame is immovable with respect to the observer. It can be used to prove that the four point-like features correspond to four locations (compare [155]). From Axiom 7, it follows that all foci on the same point-like features constitute the same reference frame. This assures there is only one configuration of point-like features that is denoted by the predicate *RefFrame*.

Calibration of geometric reference frames. Furthermore, humans have an equilibrium sense for detecting whether their body is in upright direction, and are also able to detect *verticality* of directions using any kind of physical perpendicular. Such senses may be used in order to identify directions in the environment and to calibrate reference frame with them.

This observation primitive is called $VertAln$. It describes a symmetric and transitive step. Assuming collinearity and parallelism for this primitive would be an oversimplification, because *gravity lines through the earth's body are not straight lines* (compare Figure 5.13).

[25]Orthogonality can be proved along the following lines (compare Fig. 5.11): a^*, s, a_1 lie on distinct point-like features on a line. Thus angle a^*, s, a_2 must be supplementary to angle a_1, s, a_2. Since segment $a^* a_2$ is congruent to $a_1 a_2$, and the angle sides must also be congruent by construction, triangles a^*, s, a_2 and a_1, s, a_2 must be congruent, too. Thus the supplementary angles must be congruent. The intended result now follows from the fact that congruent supplementary angles are always right angles.

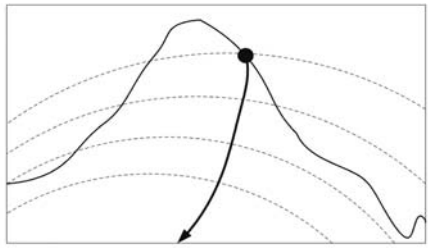


Figure 5.13: A curved gravity line and equipotential surfaces.

Axiom 10 *Symmetry and transitivity:*
$VertAln(x, y) \rightarrow VertAln(y, x) \wedge \forall z.(VertAln(y, z) \rightarrow VertAln(x, z))$

5.1.6. Identifying actions, intentions and processes

Similar to the recognition of objects, the problem of cognizing human acts was traditionally approached not from a *perceptual*, but from a *sensational/cognitivist* viewpoint. It was held that human reason not only extracts objects from an undifferentiated flow of primitive sensations, but also that actions are cognized on the basis of such extracted information. This means motor control and cognition was assumed to be separate and cascaded in a fixed series *sensation* $\Rightarrow$ *cognition* $\Rightarrow$ *motoract*. Instead, George Herbert Mead [112] argued for a tight feedback connection between action and perception one hundred years ago. It has only recently been discovered that the traditional view is in fact untenable.

Empirical evidence. Rizzolatti [145] discovered that motor acts and perception are *neurologically inseparable* to a large extent. Motor competences are actually a predisposition for perception as much as the other way around. Rizzolatti studied so-called *mirror neurons* in the premotor area of the cerebral cortex of apes[26]. He recognized that these specific cells fire in response to some *general type of action*, such as grasping for some object, regardless of whether this action was actually performed by the observer or by another person [145][27].

Premotor region 'F5' in the rear frontal cortex is more generally said to encode a *dictionary of actions*, where some neurons encode the general goal (intention), and some of them the way and order of elementary performances. In contrast, higher-level reasoning about actions (located in the prefrontal cortex) consists in *selecting among actions and action opportunities* for putting them into action. Rizzolatti's main point is that actions are "offered" to the mind by an integrated

[26]More specifically, in the premotor area 'F5' in the frontal lobe.

[27]Additionally, other neuron types in F5 fire also if the associated object was only visually present. This *action opportunity* aspect of so called *canonical neurons* is captured by the term *affordance* [145, Chapter 2] and is treated in the next subsection.

neuronal visuomotor function [145] in terms of F5 neurons. This corresponds to the view that this function is pre-conceptual and not based on reflective knowledge. In the human brain, there could exist an analogy to F5 in terms of the Brodmann area 44 [145]. The breadth of entries in this dictionary varies with individual motor experiences and seems more differentiated in the human brain than in the one of apes.

Mirror neurons fire if and only if some type of act is perceivably performed by someone. This may be the observer himself, or the act may only be partially visible, or only audible. Mirror neurons in the human brain fire even if performance is intransitive, i.e. remains without result, or is only simulated [145, Chapter 5]. They do not fire if the objects involved are the only aspects present in a scene (so they do not encode affordances) [145, Chapter 4]. They also fire for partial performances of complex synthetic acts. In this way, the underlying *intentions* of complex acts (even ambiguous ones) may be anticipated with ease. In general, we can assume that mirror neurons- or some effectively equivalent mechanisms- provide a pre-conceptual basis for constituting a *stable, actor- and modus- invariant space of actions*[28].

Observation predicates for actions and perdurants. We can safely assume that there is an autonomous perceptual apparatus for actions. But what kind of predications are enabled by action percepts and how can they be encoded?

The answer to this question seems less obvious than comparable ones in this chapter. If we perceive an act, we not only detect its *presence at a spot*, but also a continuous sequence of motor events and a transfer of objects involved, which represent essential information. Object recognition, in contrast, seems temporally more coarse: it detects only whether one and the same object is present at every time instant. This difference seems a perceptual analogue of the ontological distinction between *endurants* and *perdurants*, i.e. between things in time and events extended through time[29]. Furthermore, identity criteria for endurants and perdurants seem closely interwoven. The identity of objects seems as much ontologically prior to processes as the other way around, as Galton argued [46].

I take surface identification once more to be fundamental, but suggest that the perception of actions adds some irreducible information. Actions are therefore not ontologically secondary and not reducible to spatio-temporal configurations

[28]Recently, criticism has been raised against Rizzolatti's assumption that mirror neurons are *the* biological basis for the understanding of actions [66]. Action understanding seems to involve more than only the mirror neuron system, which is focused on its motor aspects. Nevertheless, mirror neurons undoubtedly realize a close connection between action perception and motor response [66]. Hickok assumes they allow for selecting among actions. My assumption is only that some pre-conceptual sensori-motor mechanism for fast detecting and selecting of actions must exist.

[29]Compare DOLCE [109] for an applied perspective on this distinction, and Johanna Seibt [167] for a different but interesting philosophical perspective.

of objects (in particular, not reducible to movements). I propose that action predications *follow surfaces or features as long as they participate in an action.* The predication of an action simply means to relate foci of attention that are focused on surfaces or features involved in the same action. The presence of this action is indicated by a Gestalt-like mechanism, e.g. mirror neurons. The directedness of actions can be captured by snapshots of involved objects and their spatio-temporal order. In order to capture this aspect, we simply need some binary relation expressing action equivalence of foci.

Furthermore, the neuron architecture allows to distinguish participants. The predicate therefore needs also to register *the roles that different participants play in an action.* The proposed perceptual relation has a larger arity than an equivalence relation, since I use relational order to indicate these roles. I simply assume that the first two digits in this relation are always focused on the initiator of the action, i.e., on the *actor.* They also express action equivalence among moments. Note that I assume this relation to be constitutive for the notion of a person, i.e. an intentional human being. Other roles may then be introduced by further digits of the appropriate arity. In the proposal below, I assume a ternary predicate for 2-*role actions* including an actor and some manipulated object.

Kinds of 2-role actions can be introduced by substituting some symbol, e.g. *Drink*, in the position X in the following Axiom Schema for the action predicate DoX. $(z)DoDrink(x, y)$, for example, *means that the actor in focus at moment x continues at moment y to perform the action*, i.e. to drink the substance in focus at moment z. Note that while the actor in focus stays one and the same during the action, the auxiliary surface or feature at z may change due to Axiom Schema 8. The variable X_i is thereby assumed to range over all feature types:

Axiom Schema 6 *Symmetry of DoX:*
$(z)DoX(x, y) \rightarrow (z)DoX(y, x)$

Axiom Schema 7 *Transitivity of DoX:*
$(z)DoX(x, y) \wedge (z)DoX(y, v) \rightarrow (z)DoX(x, v)$

Axiom Schema 8 *DoX is predicated on (body-) surfaces or features:*
$(z)DoX(x, y) \rightarrow SC(x, y) \wedge (\exists u. (SC(u, z) \vee \bigvee_{i=1}^{n} \exists r.(r)X_i F(u, z)))$

I assume that a certain *action type* substituted for X has a fixed arity, since actions have a fixed list of necessary participants, and so there may be actually a list of references instead of just z. The relation can be used to unify a whole action event, since the predication ceases precisely when the action stops (i.e. when the mirror neuron stops firing). A *single action event* is given by a whole $Whole^{F}_{DoX^*}$ with respect to the induced relation $DoX^*(x, y) \leftrightarrow \exists z.(z)DoX(x, y)$. Note that since the relation is transitive and symmetric, action events of the same

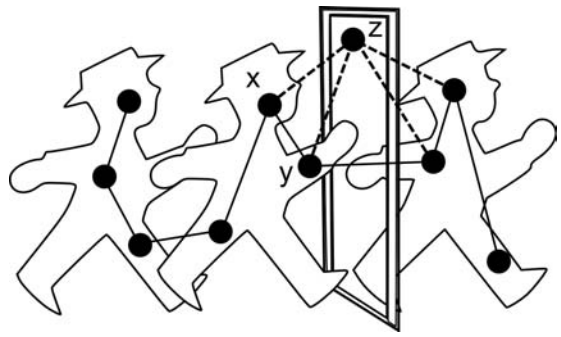


Figure 5.14: Predicating a walk through a door by $(z)DoWalkThrough(x, y)$.

type cannot overlap, i.e. they cannot be performed at the same points in space and time.

The object o *participates as an actor* in an action e if the respective wholes overlap. Remember that this means that the respective surface in focus at the moment of overlap belongs to the actor's body. Similarly, participation can be defined for other involved objects:

Definition 16 *Actor:*
$Actor(o, e) \leftrightarrow Whole^F_{DoX^*}(e) \wedge Whole^F_{SC}(o) \wedge$
$(\exists x, y, z.x \in o \wedge x \in e \wedge (z)DoX(x, y))$

Definition 17 *Participating in Action:*
$Part(o, e) \leftrightarrow Whole^F_{DoX^*}(e) \wedge \big(Whole^F_{SC}(o) \vee Whole^F_{XF^*}(o)\big) \wedge$
$(\exists x, y, z.x \in e \wedge z \in o \wedge (z)DoX(x, y))$

In order to illustrate this apparatus with a simple example, suppose we observe a person going through a door (compare Figure 5.14). The door o is an open feature $Part(o, e)$ involved in the action e of the actor a, $Actor(a, e)$. This observation could be made by the predicate $(z)DoWalkThrough(x, y)$, where $Whole^F_{DoWalkThrough^*}(e)$.

In practical situations, actions are often analyzable in terms of constituting action parts. In the simplest case, this may just mean that action parts occur one after another and share some actor or participating object. Complex actions in this case may be definable based on temporal order and object identification. Alternatively, one may think about a perceivable parthood relation among them (compare [167]).

The subjectless, functional individuation of an event type proposed here has much in common with Seibt's idea of a *general process* [167]: "it is snowing (now here)" is predicated in a similar way as "he is drinking (now here)". In all these cases, the assertion of the presence of a process is not a consequence of the presence of objects. The introduction of similar primitives for movements $MoveX$ and other process types therefore suggests itself. The particular roles of objects involved in these cases needs to be spelled out.

5.1.7. Identifying affordances by simulations

Affordance is one of the key concepts in ecological psychology. Affordances capture the functional aspect of objects in an observer's environment as well as an observer's opportunities for actions [148]. As Gibson puts it:

> "The *affordances* of the environment are what it *offers* the animal, what it *provides* or *furnishes*, whether for good or ill. [...] I mean by it something that refers to both the environment and the animal [...]." [50, page 127, emphasis in original]

An observer in this view is not only perceiving but also (potentially) acting. Gibson's own examples of affordances include *action affordances* like climb-ability (walls), catch-ability (balls), eat-ability, mail-ability (postbox), but also so-called *happening affordances* like getting burned (by fire) or falling off (a cliff) (compare [150]).

Empirical evidence. Gibson assumed that affordances can be perceived *intersubjectively*, while being *observer or actor relative* [50, page 129]. He held that affordances imply the same action potentials for the observer as well as for other actors [50, page 141]. He also claimed that they are an *independent* category of perception, prior to the perception of qualities [50, page 134].

This is supported by recent cognitive studies which emphasize the role of perceived intentionality for human cognition [188]. In the cerebral cortex of the ape, in the same frontal premotor area where mirror neurons are located, researchers also found so called *canonical cells* [145, Chapter 2] with astonishing visuomotor properties. Just as mirror neurons, they fire selectively in the context of a certain type of action, such as grasping some cup. But they do so regardless of whether the act is actually performed by someone else, by the observer himself, or whether the *visual object configuration* depicts only the opportunity to act. The reaction to visual aspects is thereby highly selective and biased towards motor relevant properties.

Rizzolatti interprets this result in the way that the respective motor region encodes *potential motor acts* [145, Chapter 2]. Since these are genuine but virtual motor representations, they can be easily understood by the observer in terms of his or her own experience. Furthermore, they can serve to categorize objects. The objects, in turn, which can be recognized also by an independent mechanism, serve as *virtual poles of action*.

Most ecological psychologists, including Gibson, have additionally claimed that affordances are *perceived directly*, i.e. without interference of cognition, and without perception of other properties [150]. However such claims seem doubtful because perception is often influenced by cognition [41].

Affordances as simulated actions. How can we make use of the notion of affordance for grounding purposes, given these empirical findings? There is a large amount of literature on affordances in ecological psychology, but most of it seems

strangely biased regarding the explanation of inter-subjectivity. Many authors attempted to objectify the notion by metaphysical claims about the nature of affordances beyond perception [190, 178]. Regarding our purpose, their suggestions raise problems.

Viewing affordances as properties of things in the environment [190] seems problematic, because they are also constituted by properties of a particular agent: Stairs are climbable only with respect to an agent's leg length (compare Warren's experiments [201]). Treating affordances as combined qualities of environments and actors (as proposed in [178]), which seems to work in the staircase example (by relating leg length and riser height), is also problematic. Take, for example, the *traversability* of a road. A road is traversable with respect to a complex interaction of an agent's crossing and the behavior of car drivers. But traversability is not a combination of a property of the agent with a property of the environment. Rather, it is an interplay of prospective events which is not reducible to any combination of properties [124]. I follow Scarantino [150] in that affordances always involve an observer's reaction. I propose to conceive of them literally as *perceivable potential events*.

But how are potential events perceived? One possible explanation is that perceived affordances are the result of *perceptual simulations*. These were proposed by Barsalou [5] in order to state that human perception and cognition are closely interlinked on the basis of perceptual simulators. They allow humans to imagine and reconstruct formerly perceived sensori-motor patterns of objects, e.g. cars, in new situations, and to reason with them. I suggest to apply this idea to affordances, saying that if pedestrians perceive the affordance of crossing a road, they do so by successfully simulating a crossing action in a given perceived scene. Perceived affordances can be "acted on", i.e., they are a necessary input to human actions, as proposed by Ortmann and Kuhn [123].

Many affordances have a social aspect, in the sense that they involve the interpretation of signs. A prominent example for a so-called *social affordance* [189] is a postbox that affords sending letters. The postbox physically only affords dropping letters (or other similarly shaped objects) through a slot. However, in the social environment that uses the appearance of boxes as conventional signs (blue in the USA, red in the UK, yellow in Germany), this box affords sending letters if the letters are properly labeled and postpaid. Since a simulative account of affordances does not exclude cognition of signs, social affordances are compatible with this approach.

Observation predicates. If we perceive an affordance, we perceive a virtual motor act. The predication of this act in a scene therefore closely resembles the predication of an action, except that the performance does not entail an acting body. Instead, the virtual actor is rather a simulated Gestalt and not a body, and may therefore "transcend" surfaces like a ghost. It may e.g. go through a closed door if the observer expects it can be opened.

Figure 5.15: A chair observed at u affords sitting at x and y to a person observed at z. This can be expressed by the 2-role predicate $(z, u)\mathit{AffordsSitting}(x, y)$.

In the simplest case, actor and participating objects are unknown to the observer, they are "dummies" of a pre-conceptual Gestalt simulation, so to speak. Since participants are not identifiable, the roles in this act are not "casted" and cannot be distinguished. It is only the presence of the affordance to be registered. For example, I may observe that a room is traversable for someone who is not further specified. These observation predicates may be called 0-role affordances. In other cases, participating objects are perceived or known, but are not acting in the scene. So the roles are casted and filled with distinguishable bodies. For example, I may observe that a chair is small enough for my son to sit on. The default role, the role of the actor in this simulated action is filled by my son.

The Axiom Schema for a 2-role affordance looks like this, where z refers to an existing person, u to some object, x and y are any foci of attention, and X is a meta-variable for any action type that may also appear in DoX:

Axiom Schema 9 *Symmetry of AffordsX:*
$(z, u)\mathit{AffordsX}(x, y) \rightarrow (z, u)\mathit{AffordsX}(y, x)$

Axiom Schema 10 *Transitivity of AffordsX:*
$(z, u)\mathit{AffordsX}(x, y) \wedge (z, u)\mathit{AffordsX}(y, v) \rightarrow (z, u)\mathit{AffordsX}(x, v)$

In Figure 5.15, the foci x and y afford a person observed at z to sit on a chair observed at u. Similar to action predications, the equivalence of foci x and y in this predicate expresses that the performance of the very same act at focus x is continued at focus y. The spatio-temporal property of an affordance event therefore indicates *when and where an action of a certain type can be performed.* This accounts for the fact that affordances are concrete percepts, and therefore do not have unconfined validity. For example, a door may be locked and loose its affordance of being opened.

Note that I consider affordance predicates – like predicates for performed actions – to be transitive. This means I assume that two different affordances of the same type cannot be simulated at the same point in space and time, i.e., cannot overlap. This may be too restrictive an assumption and then could be dropped.

The role schema can be extended in the same way as for actions. In contrast to DoX, I do not restrict the domain of affordances and their role fillers to be bodies or surfaces, because affordances can be simulated everywhere and with reference to immaterial things[30]. But note that since the equivalence of x and y in this predicate cannot itself indicate the filled actor role z, an n-role affordance is $(n+2)$-ary, whereas the corresponding n-role action is $(n+1)$-ary. If the same action type is used in affordance predicates of different arity (this was excluded for actions, see p. 105), this should mean by convention that the missing digits correspond to roles that are not casted, as proposed above.

These affordance predications play a central role in my methodology. They are used to construct a central category of the meaningful environment, namely a medium (compare Section 6.3).

5.1.8. Identifying colors, shapes, objects and matter

Surfaces are at the heart of perception. They are bearers of surface qualities like shape and color that allow us to identify individual objects over long periods of absence. Also, the texture of surfaces allows us to identify kinds of matter [50]. From these ideas, only color is discussed in more depth here as an example for a similarity predication.

Empirical evidence for color Gestalt mechanism. In the study of Kay and McDaniel [78], the authors suggested that there are 6 different types of color-sensitive opponent cells for detecting the primary colors blue, yellow, red, green, black and white. Each of these fire with a certain probability that can be modeled by a fuzzy membership function of measured wavelengths and intensities. By using fuzzy set intersection, the non-primary colors orange, purple, pink, brown and gray can be derived, whereas fuzzy set union could account for broader color categories as cool and warm. Using this mechanism, the 11 so called *focal colors* and their unions can be reconstructed, which are known to be of universal significance as basic color terms in every human language. Even though the theory has been criticized and modified with regards to its claims about natural language color terms, the basic insight, that a universal neurophysiological mechanism of primary color detection exists from which most color terms of most natural languages can be constructed, remains valid [77].

[30]For example, in Chapter 6, I will introduce the essential predicate *AffordsSeeing*, which has a visible point as a role filler in an observed seeing affordance.

Nevertheless, color perception is subject to a non-trivial Gestalt mechanism that cannot be reduced to wavelengths and intensities. As I mentioned in Section 3.2, the perception of white in a dark and a bright context varies dramatically in terms of wavelengths. It still is recognized as the same color by way of a Gestalt invariance mechanism that needs to take this context into account.

Observation predicate for color. The uncertainty of color predication is apparent because primary colors are prototypes and the color boundaries are fuzzy. Colors are the classical example for a graded category [90]. The fuzziness of this mechanism can be captured by a *similarity predication* (compare Section 4.2), i.e. a symmetric but non-transitive binary relation that asserts that focus x is similar in color to focus y:

Axiom 11 *Symmetry of* $=_{Color}$*:*
$x =_{Color} y \rightarrow y =_{Color} x$

Since wholes of this predicate are overlapping, colors may be constructed as a $QualeWhole^{F}_{=_{Color}}$, a maximal non-dissected whole part[31]. This class represents an atomic region in a color space, because members cannot be distinguished anymore (compare [131]). Other classes may be constructed based on unions of these sets. In general, it depends on our ability to make color distinctions based this Gestalt mechanism, and on the number of compared color samples, how many colors may be constructed that way.

Shapes and objects. The construction of object-based qualities like shape is a more complex case which seems to involve the construction of body parts, as well as perceiving their geometric configuration. I stick to the principle idea of Marr and Nishihara [106], that shape recognition involves a process of retrieving, storing, and comparing complex shape descriptions of bodies. These could e.g. be descriptions of oriented cylindrical body parts, consisting in the lengths of such parts and the object centered orientation angles between their natural axes [106]. The underlying operation may be physically realized as a pre-conceptual neurological mechanism, but could also be consciously reconstructed by measurements. In either case, if such shape descriptions are compared with the help of tolerance ranges, this automatically induces similarity classes as shapes on them, similar to colors. Similarity of shapes is uncertain and non-transitive, since it depends on the shape resolution and tolerance levels which are arbitrary to a certain extent (compare the discussion in [106]).

5.1.9. Overview of proposed observation predicates

In this section, I have reviewed empirical evidence for certain kinds of perceptual operations. I have also made proposals for corresponding formal observation

[31] Compare Section 4.2, and also Carnap's treatment in [25].

predicates which play the role of logical primitives, and motivated some of their axioms. In the subsequent chapters, I will formulate a grounded FOL reference theory based on these primitives. Proposals were made based on empirical evidence about underlying perceptual operations. Perceptual operations are predications which can be performed in the ISR domain and which depend on some kind of pre-conceptual Gestalt mechanism.

Name	Formal symbol	introduced in
focus of attention	F	Section 5.1.1
temporal order	$\leq_T$	Section 5.1.1
surface connection	SC	Section 5.1.3
feature	XF	Section 5.1.4
point-like feature	PF	Section 5.1.4
equal length	$=_L$	Section 5.1.5
spatial order	OnL	Section 5.1.5
verticality	$VertAln$	Section 5.1.5
action	DoX	Section 5.1.6
affordance	$AffordsX$	Section 5.1.7
similar color	$=_{Color}$	Section 5.1.8

Table 5.1: Inventory of observation predicates. Predicates including an X are predicate schemas that require substitution of X with some type.

There is a surprisingly rich inventory of such predications (see Table 5.1). Besides the traditionally considered forms of experience, such as form, color and shape, they span diverse domains of cognition, ranging from temporal experience, over surfaces and bodies, features (like edges, holes and point-like features), geometric comparisons and localizations, actions, intentions, processes and affordances. In other words, these predications provide the elementary equipment for referring to the most basic things we are able to talk about. What is new about this compilation is that most elements were traditionally considered domains of conceptual knowledge, not of pre-conceptual mechanisms. The inventory illustrates that reference is not a one-way street, and that the roots of reference, which need to be traced outside language, are delicate and manifold.

5.2. A grounding method, its metalogical properties and application scope

It remains to give a synopsis of the meta-theory of grounding proposed in the last two chapters. In particular, I discuss *reference theories*, their metalogical properties and scope of application.

Meta-theory of grounding. The meta-theory supplies logical vocabulary for grounded theories in terms of observation predicates (perceptual sources of grounding). It also describes how these predicates should be interpreted into the private

sensory-motor apparatus, and how their conventional meanings can be established inter-subjectively based on joint attention, for example by ostension. Note that the observation predicates proposed in the last section are preliminary suggestions. Their choice not only depends on cognitive and perceptual research results to come, but also on the respective information community needs. In principle, every inter-subjective perceptual competence that one can refer to in the space around the body can serve as a perceptual source of grounding. The list of observation predicates is therefore neither complete nor definite. The meta-theory also suggests a constructive observation language in FOL that can be used to guide the construction of other ideas based on these predicates (constructive source of grounding). Following Carnap and others, I have proposed a certain form of a finite class, namely a unified whole, as a useful construction. The usefulness of this idea will be demonstrated in the next chapters. Other constructive guidance principles remain to be illustrated.

Grounded reference theories. As proposed in the introduction (Section 1.2), the meta-theory should be used to formulate particular grounded reference theories. These are fully formalized FOL theories that include observation predicates as non-logical primitives of their vocabulary and serve to reconstruct particular domains of experience. As an example, I will propose a formalization of Gibson's meaningful environment and its experiential geometry in the next chapter. This theory is intended as a foundation for other reference theories.

Reference theories have a *conventionally established interpretation* into conscious mental domains that are a product of human operations. In particular, they describe the domain of attentional moments, which is a product of perceptual operations such as focusing attention and performing predications. They therefore have an extensional formal semantics as well as an operational meaning. But note that this *operational interpretation* of a reference theory *is independent of a particular formalization.* This means that while the interpretation of an observation predicate in terms of perceptual operation results is conventionally established among observers, its formal description by axioms remains an open and flexible business. This is helpful since different purposes may need different formal depths and inference machinery, while the interpretation remains fixed. It also seems inevitable because reference theories are common-sense empirical theories which need to be maintained and defended.

Consistency and incompleteness of reference theories. A formal theory is *consistent* if it does not contain any contradictions. This is just the case if it has a model, i.e. if there exists an interpretation of the theory which satisfies all axioms. Therefore we can prove a theory to be consistent if we can find a model.

A first-order theory is said to be *complete* if it is a maximal consistent set of sentences, i.e., if it contains either every sentence that can be formulated in its language, or its negation. Since all models of a theory satisfy its sentences,

an equivalent requirement is that any sentence of the language must either be satisfied by all or none of its models [186]. In this case, all models satisfy the very same sentences, i.e. they are *elementary equivalent*. Therefore, we can prove a theory to be incomplete if we find two models that are not elementary equivalent, i.e. if there is a sentence true in one but not in the other model.

Because intended interpretations of reference theories in terms of memory provide such a model, they should be consistent. For the reference theory which is outlined in the following chapters (compare the list of axioms in Appendix 8.4), I have proved that it has two simple models that are not elementary equivalent, and therefore the theory must be consistent but incomplete. The poof can be found in Appendix 8.3. This is probably the case for most ontological theories. Bennett [9] has argued that ontologies should be intended to be complete, because they – in a sense – then explicitly say everything expressible in their language. An argument against this - apart from the known absolute limitations of completeness - is that incomplete ontological theories may reflect an intended lack of empirical knowledge, which deliberately lets the truth of some sentences be undecided.

Finiteness and non-categoricity of reference theories. A formal theory is called *finite* if all its models are finite structures, i.e., have a finite universe. Reference theories are assumed to be finite because humans can attend only to a finite excerpt of their experience and can perform only finitely many constructions. Since the domain of interpretation D of every reference theory consists of F and a finite list of reified entities by Axiom Schemas 1 and 2, D must be finite, too. This is the reason why in Chapter 6, I propose a finite version of experiential geometry, as an explicit construction out of a potentially continuous geometric experience.

A useful aspect of finiteness is that we can use *recursive definitions* to construct things. In fact, finiteness of reference theories just mirrors their constructibility in terms of recursive rules, as discussed in Section 4.1.2. For example, even though the *transitive closure* of a binary relation is not in general first-order definable[32], it is possible to construct it in terms of an exhaustive finite chaining. The *formal semantics* of a recursive formula is normally given by a least fixpoint [33, page 220], however the latter transcends the ordinary semantics of FOL. But if it is possible to know the finite cardinality $n = |D|$ of the domain D in advance, the corresponding finite transitive closure R^{TC} of a relation, denoted by wild card R, can be defined in ordinary FOL as follows:

Definition 18 *Finite transitive closure of a binary relation R:*
$R^{TC}(x, y) \leftrightarrow (R(x, y) \vee \bigvee_{i=1}^{n}(\exists z_1, ..., z_i.R(x, z_1) \wedge \ldots \wedge R(z_i, y)))$

We will use this definitional schema in the following chapters in order to construct finite chains in an exhaustive manner. We thereby do not assume more than that

[32]This is because a transitive closure is defined as the *minimal* transitive super relation of a relation, and this minimality condition cannot be formalized in FOL.

an observer can count his observation records and his acts of logical reification before constructing transitive chains.

A theory is called categorical if all its models are isomorphic (one-to-one correspondent) among each other. This basically means that models cannot be distinguished by their structure. First-order theories with finite domains can be categorical[33]. But since categoricity entails completeness [4], the proposed reference theory is not categorical.

Bennett [9] has argued that one practical reason for non-categoricity of ontologies may be *contingency*, i.e., the ability to generalize over structurally different worlds. The reference theories proposed here are a good example for this. They need to have models that are not one-to-one correspondent, because each intended interpretation of a reference theory is a *different state of memory of an observer* (compare Section 4.2). If the reference theory was categorical, it could express only one such state. This would not be very useful for the purpose of referencing, which occurs in very different states of perceptual memory. It follows that reference theories are also conceptually incomplete [9].

Application scope. A grounded reference theory can be used to describe a perceptual state of memory of an observer, including results of joint predications in the space of immediate perception (ISR). It is based on a conventionally established observation language with an operational semantics.

From the viewpoint of information semantics, grounded reference theories provide a means to say very clearly how a certain term or data item should be interpreted into observations. This is needed for comparing data descriptions on the most general interpersonal semantic level that exists, which is our competence of joint perceptual predication. I expect that all domains of geographic information can be described in this way.

Reference theories thereby take the role of *top-level ontologies* that provide a background for comparing or designing application ontologies in particular representation languages [57, 86], such as RDF[34]. But in contrast to known top-level ontologies, like DOLCE [109], reference theories do not abstract away from observation procedures and thereby allow to reconstruct data by observation. This can also help to resolve the grounding problem of ontologies. I will discuss areas of application in the following chapters.

[33]Whereas, because of the Löwenheim-Skolem theorem, FOL theories cannot distinguish non-denumerable from denumerable models, and thus infinite theories are never categorical.

[34]See `http://www.w3.org/RDF/`.

Part III

Grounding geodata in the meaningful environment

Chapter 6

Constructing the meaningful environment

The doctrine that we could not perceive the world around us
unless we already had the concept of space is nonsense.
It is quite the other way around:
We could not conceive of empty space
unless we could see the ground under our feet and the sky above.

— J.J. Gibson [50]

So far, I have argued for an attentional apparatus for predication and reference. It allows humans to join their attention on shared perceptual Gestalt phenomena in their space around the body (Chapter 3), and thus to solve the problem of inter-subjective symbol reference. I have also suggested constructive means of data grounding in terms of a technical observation language, a language that allows to talk about acts of focusing attention as basic information items. It allows to introduce formal observation predicates as primitives and accounts for the freedom of reification of abstract information items (Chapter 4). I have furthermore discussed perceptual sources of data grounding, including justifiable suggestions for observation predicates (Chapter 5). In this chapter, I will demonstrate how these sources can be fruitfully combined in order to reconstruct major domains of experience underlying geographic information. I formalize useful notions of information that describe geographic data and prove some of their intuitively expectable properties. In the subsequent chapter, I will discuss their application to navigation networks in greater detail.

One may wonder why I run this demonstration under the heading of Gibson's meaningful environment (compare Section 5.1.2). The reason is that I view Gibson's environmental terms among the most essential constructions for grounding geodata. However, even though they are basic notions to describe information, I guess they do not have to be taken as primitive. The information categories

discussed here comprise also abstract things, such as geometry and units of measurements, which at first sight might not be considered part of ecological psychology. Gibson's treatment does not distinguish between primitive and complex notions, and it does not consider constructions [50]. For a formal treatment, this distinction is crucial, and for grounding purposes, it is indispensable. I agree with Gibson's cited statement above. However, I consider not only space to be on a higher level of abstraction, but also some of his own terms.

I start with an example that illustrates the basic idea of how things and their qualities in this meaningful environment can be constructed based on attentional steps (taken from [158]).

The blind person in a closed room. How does a blind man perceive the geometric qualities of a closed room? Standing inside the room, he can rely on his tactile and hearing sensors to detect its surface qualities. Because he knows his body takes some of the space of the room, the room must be higher than his body. If he can turn around where he stands, he knows that a roughly cylindrical space is free and part of the room. By taking a step forward, he concludes that an elongated "corridor" is free and part of the room. Because he can repeat steps of the same length into the same direction, he can step diametrically through the room and even measure one of its diameters[1]. His last step may be shortened, because his foot bumps into the wall. He thus detected an inner surface of the room. If he continues summing up paths through the room, he can individuate the whole room by its horizontal extent.

6.1. Constructing a geometry in the meaningful environment

As I argued in Section 5.1.5, in order to construct experiential geometry, observers can draw on perceptual operations for comparing equidistance $=_L$ and collinearity OnL, as well as on some spatial reference frame in terms of point-like features. The latter aspect means that experiential geometry is *relativistic*, i.e., that it exists only relative to such a frame of reference, as was argued by Leibniz [95] and Poincaré [128]. Let us assume the observer has chosen some frame *RefFrame* in terms of four point-like features, as described in Section 5.1.5, so that locations can be constructed with respect to it.

In the following, I will discuss an axiomatization of a *finitist relativistic geometry*[2] which may be constructed on this basis. The axiomatization largely follows the newer Tarskian system given in [162], but deviates in being constructive, in allowing for finite models, and in not presuming the existence of a domain of

[1] If you doubt whether blind persons are in fact able to do this, you should visit the *Unsichtbar* (`http://www.unsicht-bar.com/`), a restaurant for dining in absolute darkness, and ask the blind waiters how they manage to find your table.

[2] The notion "finitist" refers to the philosophical standpoints of finitism and intuitonism.

points. It abandons infinity axioms and universal-existential claims, substituting them with constructive steps, i.e., with explicit existential quantifications (Section 5.1.5). After proving some simple and intuitive geometric facts, I illustrate the constructive idea by *constructive Theorem* 12 about segment reflections. This demonstrates that relativistic finitist FOL geometries are able to capture how observers can construct geometry in their environment. An alternative way is to describe geometric constructions not by an axiomatic theory, but more explicitly by recursive rules [100], as was done in [155], or by dialogic formalisms [100][3].

In this section, I will use same letters with apostrophe, e.g. z, z', z'', in order to denote attentional moments which are focused on the same locus with respect to this reference frame, i.e., in order to abbreviate location equivalence $z =_{Ref} z' =_{Ref} z''$ (compare Section 5.1.5 and Definition 15 and Axiom 6). This is for convenient readability of the following axioms. Quantified axioms with free apostrophe variables z', as e.g. $\forall z.\Phi(z, z')$, simply stand for quantified implications of the form $\forall z, z'.z =_{Ref} z' \rightarrow \Phi(z, z')$. Furthermore, all free variables are assumed to range over the domain D, consisting of attentional moments F and constructed reifications Rfc.

Loci are equivalence classes of location equivalent foci of attention (compare Section 5.1.5). Loci are the "points" of our constructed geometry, on which attention can be focused. In the following, I will sometimes speak of a *point in focus*, instead of a locus in focus. Note that this manner of speaking does not mean to sneak in points as primitive notions, since all I really talk about are foci of attention. One of the challenges will be to show that *relativistic* geometry, which is based on foci of attention, not locations, behaves neutral with respect to foci on the same location, as expected.

For constructive reasons, loci of attention imply a reference frame (compare Definition 15 and Axiom 6) that allows to construct coordinate systems in just three dimensions. Its existence is asserted by the following *lower dimension axiom*. This axiom restricts the number of dimensions to be at least three by requiring the existence of an appropriate reference frame[4].

Axiom 12 *Lower dimension = 3:*
$\exists s, a_1, a_2, a_3.RefFrame(s, a_1, a_2, a_3)$

First, I will discuss axioms for equidistance $=_L$. Equidistant steps are assumed to be reversible on loci of attention. That means it is always possible to return to the same locus by taking a step forward and a step back which has the same length. Furthermore, taking a step of zero length, i.e., to step on the spot, leads to the same locus (identity axiom). And two steps of the same length as a third

[3] I consider such kinds of constructivist formalizations future work, while this thesis focuses on exemplifying the general idea using classical logic.

[4] An equivalent *upper dimension axiom* would restrict the dimensions to be at most three by denying the existence of an equivalent frame for four dimensions [187]. An aspect of this is captured by Axiom 6.

one equal each other (connectivity axiom). These axioms correspond to Tarski's axioms A1-3 in [162].

Axiom 13 *Reversibility of equidistance:*
$xy =_L y'x'$

Axiom 14 *Identity of equidistance:*
$xy =_L zz' \leftrightarrow x =_{Ref} y$

Axiom 15 *Connectivity of equidistance:*
$xy =_L zu \wedge x'y' =_L vw \rightarrow z'u' =_L v'w'$

Equidistance is neutral with respect to substituting foci of attention on equivalent loci, i.e., it is reflexive on loci:

Theorem 1 *Locus neutrality (reflexivity) of equidistance:*
$xy =_L x'y'$

PROOF Suppose $xy =_L yx \wedge x'y' =_L y'x'$ by inserting the same foci into Axiom 13. Then it follows immediately from Axiom 15 that $yx =_L y'x'$. *q.e.d.*

We can now infer that equidistance gives rise to an equivalence relation among pairs of loci:

Theorem 2 *Symmetry of equidistance:*
$ab =_L cd \rightarrow c'd' =_L a'b'$

PROOF Suppose $ab =_L cd$ by assumption of symmetry. Then $ab =_L a'b'$ by Theorem 1. Then by Axiom 15, $c'd' =_L a'b'$. *q.e.d.*

Theorem 3 *Transitivity of equidistance:*
$ab =_L cd \wedge c'd' =_L zu \rightarrow ab =_L zu$

PROOF Just apply symmetry Theorem 2 to Axiom 15. *q.e.d.*

From Axiom 13 and Theorem 3 follows also that equidistance is neutral with respect to the order of the two pairs of foci. Based on these results, one can define *line segments* as classes based on pairs of foci and "same length" as a binary equivalence relation on these classes. This will be done at the end of the next section together with the proof of metric properties of lengths.

Note that the identity Axiom 14 is a biconditional instead of a simple implication as in [162]. This assures that steps of zero length are always congruent to each other, without the need to draw on the axiom of segment construction:

Theorem 4 *Zero steps are equidistant:*
$xx' =_L yy'$

PROOF Suppose there are 2 coinciding pairs of foci x, x' and y, y'. These pairs satisfy the right hand side of Axiom 14, and since it is a biconditional, it follows that $xx' =_L yy'$. *q.e.d.*

The following axioms characterize the collinearity relation OnL. These are quite numerous compared to [162] because I have to compensate for the loss of the Axiom of Pasch, which is one of the infinity axioms in Tarski's system [187] (compare Section 5.1.5). It was used in [162] to draw many inferences for OnL. First, OnL also satisfies some identity axiom: Pointing between two co-located foci means to point at the same locus. The reflexivity axiom assures that OnL applies if two of three foci coincide, i.e., two foci are always aligned. And symmetry captures the fact that three foci on a line can be ordered in two ways. The other axioms assure that two OnL-triples with two loci in common are always ordered as one would expect.

Axiom 16 *Identity of OnL:*
$OnL(x, y, x') \rightarrow x =_{Ref} y$

Axiom 17 *Reflexivity of OnL:*
$OnL(x, y, y')$

Axiom 18 *Symmetry of OnL:*
$OnL(a, b, c) \rightarrow OnL(c, b, a)$

Axiom 19 *(Inner) Transitivity:*
$OnL(a, b, d) \wedge OnL(b', c, d') \rightarrow OnL(a', b', c')$

Axiom 20 *(Outer) Transitivity:*
$OnL(a, b, c) \wedge OnL(b', c, d') \wedge \neg b =_{Ref} c \rightarrow OnL(a', b', d') \wedge OnL(a', c', d')$

Axiom 21 *(Outer) Connectivity:*
$OnL(a, b, c) \wedge OnL(a', b', d) \wedge \neg a =_{Ref} b \rightarrow (OnL(a', c', d') \vee OnL(a', d', c'))$

OnL is neutral with respect to substitutions of foci of attention on equivalent loci:

Theorem 5 *Locus neutrality of OnL:*
$OnL(x, y, z) \rightarrow OnL(x', y', z')$

PROOF Suppose $OnL(x, y, z)$ by condition. Then $OnL(y, z, z')$ by Axiom 17. The conjunction of these statements satisfies the condition of Axiom 19, and thus $OnL(x', y', z')$. *q.e.d.*

Theorem 6 *Inferences for two OnL-triples on a line:*
$OnL(a, b, d) \wedge OnL(b', c, d') \rightarrow OnL(a', c', d')$
$OnL(a, b, c) \wedge OnL(a', c', d) \rightarrow OnL(b', c', d') \wedge OnL(a', b', d')$

PROOF First statement: Applying Axiom 19 to the condition yields $OnL(a', b', c')$. Applying Axiom 20 to $OnL(a', b', c') \wedge OnL(b', c, d')$ yields the required result $OnL(a', c', d')$. Second statement: If we symmetrically convert the condition by Axiom 18, we get the condition of Axiom 19, and thus $OnL(d', c', b')$, whose symmetrical conversion yields the required first result. The second result is obtained by applying the first statement of this Theorem instead of Axiom 19. *q.e.d.*

Now we need to add axioms governing the interaction of the two observation predicates. We need to divert from Tarski's scheme here because his axioms involve closure of the domain, as argued in the first paragraph of Section 5.1.5. This is also a part of the theory which is still preliminary and needs further work, since it remains to show which Euclidean properties exactly can be saved into finiteness. It turns out that for the requirements of this thesis, we can get along with a single further axiom.

The so called *five segment axiom* [187] allows to express length summations as well as the characterization of angles. I follow [162] in the following abbreviation and illustration of this central axiom. As shown in Figure 6.1, the axiom states that in a certain configuration of four segments, the length of a certain fifth segment is always fixed.

Definition 19 *Five Segment Configuration:*
$AFS\left(\begin{smallmatrix} a & b & c & d \\ a^* & b^* & c^* & d^* \end{smallmatrix}\right) \leftrightarrow OnL(a, b, c) \wedge$
$OnL(a^*, b^*, c^*) \wedge ab =_L a^*b^* \wedge bc =_L b^*c^* \wedge ad =_L a^*d^* \wedge bd =_L b^*d^*$

Axiom 22 *Five Segment Axiom:*
$AFS\left(\begin{smallmatrix} a & b & c & d \\ a^* & b^* & c^* & d^* \end{smallmatrix}\right) \wedge \neg a =_{Ref} b \rightarrow cd =_L c^*d^*$

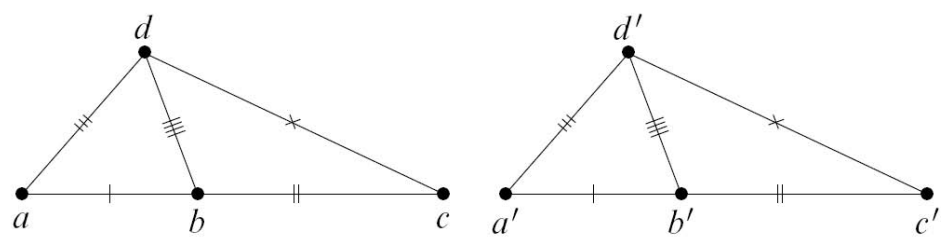


Figure 6.1: The Five Segment Axiom. The length of segment cd is fixed if a, b, c, d exhibit a five segment configuration, cf. [187].

Using this axiom, many essential theorems can be deduced [162]. For the degenerate case (the case that d coincides with a, compare Fig. 6.1), we can infer an important additive property for segments: Adding congruent segments in a line yields congruent summation segments.

Theorem 7 *Additivity of Segments:*
$OnL(a, b, c) \wedge OnL(a^*, b^*, c^*) \wedge ab =_L a^*b^* \wedge bc =_L b^*c^*$
$\rightarrow ac =_L a^*c^*$

PROOF The condition of this theorem obviously satisfies $AFS\left(\begin{smallmatrix} a & b & c & a \\ a^* & b^* & c^* & a^* \end{smallmatrix}\right)$. Distinguish two cases: For $\neg a =_{Ref} b$, the result directly follows from Axiom 22. Otherwise, by Axiom 14, it follows that $a^* =_{Ref} b^*$, and since $bc =_L b^*c^*$, so $ac =_L a^*c^*$. *q.e.d.*

We cannot use Tarski's axiom of segment construction (compare Section 5.1.5). But Axiom 22 allows to prove that any segment construction, if it is carried out, yields a unique locus. A segment is constructed by extending some segment q, a for the length of some other segment b, c up to focus x (compare Fig. 5.8).

Theorem 8 *Uniqueness of Segment Construction:*
$\neg q =_{Ref} a \rightarrow$
$(\forall x, x^*. OnL(q, a, x) \wedge OnL(q, a, x^*) \wedge ax =_L ax^* \rightarrow x =_{Ref} x^*)$

PROOF Suppose there are two foci x and x^* constructed in the way described above. From $ax =_L ax^*$, according to Theorem 7, we also get $qx =_L qx^*$ and thus the five segment configuration $AFS\left(\begin{smallmatrix} q & a & x & x \\ q & a & x & x^* \end{smallmatrix}\right)$. Thus $xx =_L xx^*$, and by Axiom 14 $x =_{Ref} x^*$. *q.e.d.*

In an analogous way, we can define an *inner five-segment configuration*, in which "inner segment" db plays the same role as dc in AFS, compare Fig. 5.8:

Definition 20 *Inner Five Segment Configuration:*
$IFS\left(\begin{smallmatrix} a & b & c & d \\ a^* & b^* & c^* & d^* \end{smallmatrix}\right) \leftrightarrow OnL(a, b, c) \wedge$
$OnL(a^*, b^*, c^*) \wedge ac =_L a^*c^* \wedge bc =_L b^*c^* \wedge ad =_L a^*d^* \wedge cd =_L c^*d^*$

We assume a corresponding Axiom for the inner five segment configuration[5].

Axiom 23 *Inner Five Segment Axiom:*
$IFS\left(\begin{smallmatrix} a & b & c & d \\ a^* & b^* & c^* & d^* \end{smallmatrix}\right) \rightarrow bd =_L b^*d^*$

We can now define some missing complementary sentences based on this Axiom:

Theorem 9 *Subtractivity of Segments:*
$OnL(a, b, c) \wedge OnL(a^*, b^*, c^*) \wedge ac =_L a^*c^* \wedge bc =_L b^*c^*$
$\rightarrow ab =_L a^*b^*$

PROOF The condition of this theorem satisfies Definition 20 with symbol d substituted with a. The result follows directly from Axiom 23. *q.e.d.*

Theorem 10 *Uniqueness of Segment Construction (Inner Form):*
$OnL(a, x, q) \wedge OnL(a, x^*, q) \wedge ax =_L ax^* \rightarrow x =_{Ref} x^*$

PROOF From $ax =_L ax^*$, according to Theorem 9, we get $qx =_L qx^*$, and thus the inner five segment configuration $IFS\left(\begin{smallmatrix} a & x & q & x \\ a & x & q & x^* \end{smallmatrix}\right)$. Thus $xx =_L xx^*$, and by Axiom 14 $x =_{Ref} x^*$. *q.e.d.*

[5]This is a theorem in Tarski's original system, since it can be proved based on the axiom of segment construction. I have decided to add it as axiom since it is fundamental and could only be proved in constructive form otherwise. This would make our system too cumbersome.

The five segment axiom also allows to prove an important property about *reflections*, namely that the reflection of a segment in a locus always yields a congruent segment. We first introduce point reflections. A reflection of a point in focus p in some other point in focus m is simply a point in focus p^* generated by doubling the segment pm (compare also Figure 6.2):

Definition 21 *Point Reflection:*
$Rf(p, m, p^*) \leftrightarrow OnL(p, m, p^*) \wedge pm =_L mp^*$

Note that from Theorem 8 it directly follows that a reflected focus p^* must coincide with any p^{**} constructed in the same way. Now we can prove that two points in focus a, b and their reflections through a point m are always congruent if they have the same distance to m, i.e., if they form a rectangle (compare Figure 6.2):

Theorem 11 *Congruence of Segments Reflected by equal distance:*
$Rf(a, m, a^*) \wedge Rf(b, m, b^*) \wedge am =_L bm \rightarrow ab =_L a^*b^*$

Proof By applying transitivity Theorem 3 to the condition above, we get $a^*m =_L bm \wedge am =_L b^*m$. Since also $ba^* =_L a^*b$ by Axiom 13, this satisfies the five segment configuration $AFS\left(\begin{smallmatrix} b & m & b^* & a^* \\ a^* & m & a & b \end{smallmatrix}\right)$. By Axiom 22, it follows that $ab =_L a^*b^*$. *q.e.d.*

We would like to prove now the same kind of segment congruence for the general case in which distances to m are not equal, i.e. in case a, c, a^*, c^* form a parallelogram (compare Figure 6.2). It turns out that we cannot do this without constructing additional "auxiliary points". This is the price we have to pay for constructive geometry: foci do not exist unless they were generated explicitly. We first have to inscribe an auxiliary rectangle into the parallelogram. So in contrast to sentence 7.13 in [162], this needs to be a *constructive* theorem which stipulates auxiliary foci. This means it needs to be of the following *existentially conditioned form* (auxiliary constructions indicated by square brackets [...]) :

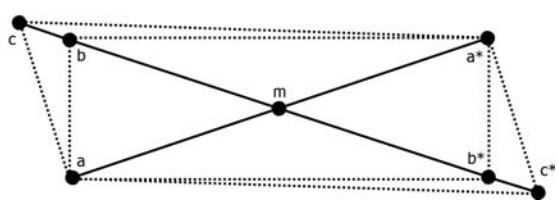


Figure 6.2: Reflection of Segments ab and ac in m always results in congruent images. In order to prove this for the general case, the parallelogram in ac, rectangular auxiliary focus b is needed.

Theorem 12 *Congruence of Reflected Segments (Constructive):*
$Rf(a, m, a^*) \wedge Rf(c, m, c^*) \wedge$
$[\exists b, b^*.am =_L bm \wedge Rf(b, m, b^*) \wedge OnL(c, b, m)] \rightarrow ac =_L a^*c^*$

PROOF Using Axiom 20 and Theorem 6, we can infer from the conditions $OnL(c, b, m) \wedge OnL(b, m, b^*) \wedge OnL(c, m, c^*)$ that $OnL(c, m, b^*) \wedge OnL(b, m, c^*)$. Now we can apply Theorem 7 to consecutive segment pairs b^*mc and bmc^*, and get $b^*c =_L c^*b$. Also, by condition we know that $bm =_L b^*m \wedge cm =_L c^*m$. We thus have a degenerated five segment configuration $AFS\left(\begin{smallmatrix} b^* & m & b & c \\ b & m & b^* & c^* \end{smallmatrix}\right)$. This delivers congruence for the two segments $bc =_L b^*c^*$, which correspond to the subtractions of parallelogram and rectangle (see Figure 6.2). By applying Theorem 11 to the rectangle we also know that $ab =_L a^*b^*$. Thus we have the non-degenerated five segment configuration $AFS\left(\begin{smallmatrix} m & b & c & a \\ m & b^* & c^* & a^* \end{smallmatrix}\right)$, which delivers the required result. *q.e.d.*

The proposed axiom set so far does not incorporate anything equivalent to the *parallel postulate* or any of its alternatives. Such an axiom would allow to infer that there is at most one parallel line through a focus not on a line. The current proposal therefore is a finite version of *absolute geometry*. It could be turned into quasi-Euclidean by adding some finite version of the parallel postulate, maybe along the lines of the third form of Euclid's axiom in [187]. Also, it remains to be shown whether this would allow to capture *all* expected properties of Euclidean space *except its infinite closure properties*.

6.2. Measuring angles and lengths in the meaningful environment

In this section, I will introduce the notions of angle and length, and I will prove that lengths have metric properties. The prove is constructive, just as Theorem 12, but also slightly more involved since we need to prove the *triangle inequality* based on angles. This will then finish my treatment of experiential geometry to cope with the example scenarios.

We begin by defining a "smaller than" order relation between segments that expresses *length comparison*. We first define two preliminary relations that are "not yet" an order. Segment ab is smaller than cd, *if there exists an expansion* ax *or a contraction* cx on the respective line that is congruent to the other pair. We distinguish these two possibilities because they reflect different ways of construction.

Definition 22 *Expansion match among segments:*
$ab \prec_L^{exp} cd \leftrightarrow \exists x. OnL(a, b, x) \wedge ax =_L cd$

Definition 23 *Contraction match among segments:*
$ab \prec_L^{con} cd \leftrightarrow \exists x. OnL(c, x, d) \wedge ab =_L cx$

Definition 24 *Proto-order among segments:*
$ab \prec_L cd \leftrightarrow (ab \prec_L^{exp} cd \vee ab \prec_L^{con} cd)$

In order to obtain a *partial order*, we additionally have to generate the transitive closure of the disjunction of these proto-relations. This is because unlike dense geometry, we lack the Axiom of Segment Construction that would ensure the existence of such extension/contraction foci for transitive pairs. Regarding transitive closure, see Definition 18.

Definition 25 *(Transitive) order among segments:*
$ab \leq_L ef \leftrightarrow ab \prec_L^{TC} ef$

That this relation is a partial order can be proved easily[6]. This is the corresponding *strict order*:

Definition 26 *Strict order among segments:*
$ab <_L cd \leftrightarrow ab \leq_L cd \wedge \neg ab =_L cd$

We can now define a quasi algebraic notion of length addition based on these notions:

Definition 27 *Addition of lengths:*
$ef + gh \leq_L ab + cd \leftrightarrow \exists e^*, g^*, h^*, a^*, c^*, d^*. OnL(e^*, g^*, h^*) \wedge OnL(a^*, c^*, d^*) \wedge$
$ab =_L a^*c^* \wedge cd =_L c^*d^* \wedge ef =_L e^*g^* \wedge gh =_L g^*h^* \wedge e^*h^* \leq_L a^*d^*$

Definition 28 *Single sums:*
$gh \leq_L ab + cd \leftrightarrow gg + gh \leq_L ab + cd$
$ef + gh \leq_L cd \leftrightarrow ef + gh \leq_L cc + cd$

Note that transitivity of this summation order follows directly from transitivity of $\leq_L$.

Second, we need to define *angles*. We first abbreviate the fact that a focus a^* lies "on the same side" of a focus b with respect to another focus a, i.e., it is on the ray starting from b in the direction of a:

Definition 29 *Being on same side:*
$a \cong_b a^* \leftrightarrow$
$\neg a =_{Ref} b \wedge \neg a^* =_{Ref} b \wedge (OnL(b, a, a^*) \vee OnL(b, a^*, a))$

Now we can define the notion of an angle, and, especially, a congruence relation between angles. Angles are taken to be corners of arbitrary triangles, i.e., they are foci triples such as abc, where b is the vertex of the angle. Congruence among angles $abc =_A def$ can be defined based on the existence of congruent triangles a^*bc^* and d^*ef^* inscribed into those angles (Figure 6.3):

Definition 30 *Congruence among Angles:*
$abc =_A def \leftrightarrow \exists a^*, c^*, d^*, f^*. a \cong_b a^* \wedge c \cong_b c^* \wedge$
$d \cong_e d^* \wedge f \cong_e f^* \wedge ba^* =_L ed^* \wedge bc^* =_L ef^* \wedge a^*c^* =_L d^*f^*$

The proof that this relation is *symmetric* and *transitive*, and that it is neutral with respect to focus order in each triangle, i.e., $abc =_A cba$, is obvious and omitted here. Based on this, we can determine whether an angle is smaller than another one. This is just the case if the bigger angle encloses another angle congruent to the smaller one.

[6]Reflexivity: Consider $ab \prec_L ab$ and just take x in Definition 24 to be b. Antisymmetry: Follows from the easily proved Theorem $OnL(a, b, x) \wedge OnL(a, x, b) \rightarrow x =_{Ref} b$.

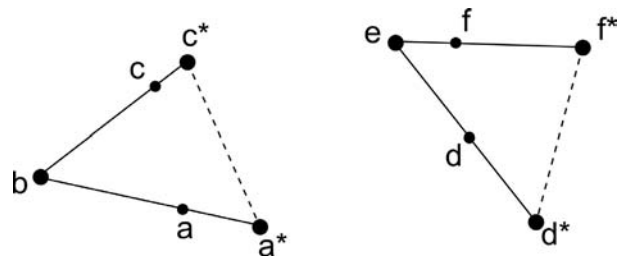


Figure 6.3: The definition of congruence among angles abc and def based on auxiliary foci a^*, c^*, d^*, f^*.

Definition 31 *Order among Angles:*
$abc \leq_A def \leftrightarrow \exists p. OnL(d, p, f) \wedge abc =_A dep$

We can now easily prove that in any isosceles triangle, the two angles opposite of b of two equilateral sides $ab =_L cb$ are equal:

Theorem 13 *Isosceles triangle theorem :*
$ab =_L bc \rightarrow bac =_A bca$

Proof Since $ba^* =_L bc$ by condition, and (trivially) $bc^* =_L ba \wedge ac =_L ca \wedge a \cong_c a \wedge b \cong_c b \wedge b \cong_a b \wedge c \cong_a c$, triangles bca and bac satisfy Definition 30. *q.e.d.*

Another simple theorem can be proved about angle relations in a constructed triangle that represents the length of a non-straight path in another triangle. It states that for a given triangle abc, if we extend the segment bc linearly to a focus c^*, such that segments cc^* and ac are congruent, then angle bc^*a needs to be smaller than angle bac^* (compare Figure 6.4):

Theorem 14 *Theorem about non-straight path triangles:*
$OnL(c^*, c, b) \wedge ac =_L cc^* \rightarrow bc^*a \leq_A c^*ab$

Proof Applying Theorem 13 to the condition $ac^* =_L cc^*$ yields $cc^*a =_A cac^*$. Since $OnL(c^*, c, b)$ by condition, we know that $c^*ac \leq_A c^*ab$ by definition 31. As mentioned above, $=_A$ is neutral with respect to focus order, $c^*ac =_A cac^*$, and it is transitive. Therefore $cc^*a \leq_A c^*ab$ by transitive combination of the last two statements. From $OnL(c^*, c, b)$ it follows that $c \cong_b c^*$ by definition 29, and therefore $cc^*a =_A bc^*a$ by Definition 30. Combining the last two statements gives the result $bc^*a \leq_A c^*ab$. *q.e.d.*

In order to prove the triangle inequality, we have to introduce an *auxiliary point configuration* ($APTI$) that allows us to reflect and compare angles in the non-straight path triangle a, b, c^*. The construction of this configuration and the proof is discussed in detail in the Appendix 8.2.

Theorem 15 *Triangle Inequality (Constructive):*
$APTI(a, b, c) \rightarrow ab \leq_L bc + ca$

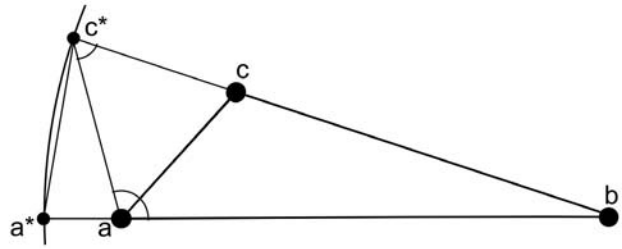


Figure 6.4: The segment c^*b reflects the length of the non-straight path bca in a triangle. The proof of the triangle inequality consists in showing that for a congruent segment ba^*, a^* must lie on or behind a with respect to b. Compare Theorems 14 and 15.

In order to finish our demonstration of metric properties of the constructed geometry, we need to introduce a length space and distance functions that map into this space. *Lengths* can be introduced as classes of equidistant pairs of foci, along the line of thought in Section 4.2. Very similar to wholes, they are nonempty maximal classes of foci pairs self-connected by equidistance. This is stated by the following definition. Together with symmetry (Theorem 2) and reversibility (Axiom 13) of equidistance, it assures that all pairs with equal length are included in this class:

Definition 32 *Classes of pairs:*
$(a, b) \in i \leftrightarrow \exists y. \{a, b\} = y \wedge y \in i$

Definition 33 *Length classes:*
$Length(i) \leftrightarrow \exists a, b.(a, b) \in i \wedge (\forall c, d.ab =_L cd \leftrightarrow (c, d) \in i)$

We also assume that such classes have been introduced as entities in terms of a finite list of number constants $0, ..., 1, ..., n$ (including zero and one) in the domain of discourse by Axiom Schema 1. It remains to introduce a (partial) distance function, fix the meaning of the constants 0 and 1 (based on the reference frame), and define an order among distances:

Definition 34 *Distance function:*
$distance(a, b) = i \leftrightarrow Length(i) \wedge (a, b) \in i$

Definition 35 *Zero length:*
$distance(a, b) = 0 \leftrightarrow a =_{Ref} b$

Definition 36 *Unit length:*
$distance(c, d) = 1 \leftrightarrow (\forall s, a_1, a_2, a_3.RefFrame(s, a_1, a_2, a_3) \rightarrow cd =_L sa)$

Definition 37 *Distance order:*
$i \leq_D j + k \leftrightarrow (\exists a, b, c, d, e, f.$
$distance(a, b) = j \wedge distance(c, d) = k \wedge distance(e, f) = i \wedge ef \leq_L ab + cd)$

The proof of metric properties of this distance function is now straightforward:

Theorem 16 *distance satisfies the metric properties:*

1. *Identity of indiscernibles:* $distance(a, b) = 0 \leftrightarrow a =_{Ref} b$
2. *Symmetry:* $distance(a, b) = distance(b, a)$
3. *Triangle Inequality (Constructive):*
   $APTI(a, b, c) \rightarrow distance(ab) \leq_D distance(bc) + distance(ca)$

PROOF The first sentence is true by definition 35. The second sentence is a direct consequence of definition 34, Axiom 33 and the reversibility Axiom 13. The third sentence is a direct consequence of Theorem 15 and Definition 37. *q.e.d.*

6.3. The individuation of meaningful things

In this section, I suggest to take Gibson literally, and at the same time, to put his direct epistemology back on its (indirect) feet. The suggestion I make is that essential categories of Gibson's *meaningful environment* can be reconstructed based on predications introduced in Section 5.1, especially affordances: *Substances* denote the things in a meaningful environment that do not afford seeing through them. *Surfaces* are located where seeing ends. More complex affordances, like *sitting* or *entering and leaving*, give rise to subcategories of substances, like *chair* and *door*. I begin by reconstructing general topological properties of Gestalts in the meaningful environment, before I focus on certain kinds of them, namely media, substances and bodies.

6.3.1. Topology of the meaningful environment

If we step through our environment, we can construct a *path* consisting of all the locations we step through (Figure 6.5). This path has special *topological properties*: it affords the *continuous transfer* of a voluminous body from one location to another. This is only possible if this path is *strongly connected*. A *strongly self-connected* path always contains a *sphere* which, when we split the path at any point, overlaps both halves of the split (compare definitions in [10, 17]). So there is a fictive surface in the middle corresponding to any cut (like "cutting in wood", see Figure 6.5), and not a "line" or a "point". This is equivalent to saying that we can move a virtual sphere through this path so that it never overlaps with foci outside. Strong connectedness can therefore be expressed based on our primitives by introducing the idea of a path in terms of a chain of overlapping spheres.

In [158], I proposed a notion of path and strong connectedness in a *continuous* version of Tarski's geometry. It was based on the idea of density. But density and continuity are not at our disposal in constructive geometry. So the question is whether there is a finite version of these notions available. We can indeed come

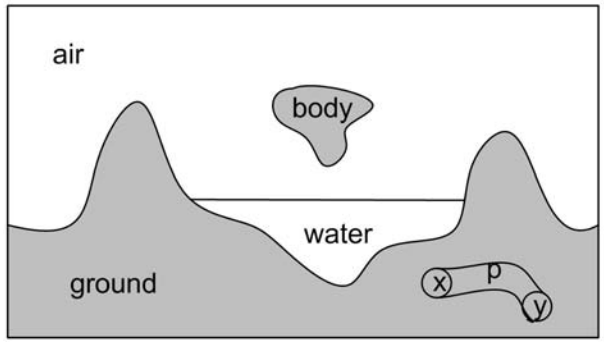


Figure 6.5: Strongly connected paths. The meaningful environment is wholly covered by strongly connected paths.

up with a notion that captures the same intuition. Finite geometry is different from dense or continuous geometry in many respects, and so my proposal diverts from [158].

Furthermore, while [158] describes a static situation, we have to deal with change. *Topological neighborhoods* are spherical subsets of foci of attention that can be used to account for connectedness. In our temporal domain of attention, these neighborhoods change as they may be occupied by substances and media at different moments. So every assertion about topological neighborhoods needs some time frame, a temporal attentional scope (denoted by pairs of foci (t_1, t_2):

Definition 38 *Temporal scope:*
$(t^*)InScope(t_1, t_2) \leftrightarrow t_1 \leq_T t^* \wedge t^* \leq_T t_2$

A topological neighborhood can be expressed by the notion of a *contained sphere*, $ContainedSphere_{\Phi}(x, y)$. This predicate asserts that the *interior* of a sphere defined by the center x and the radius xy contains - during a given temporal scope - only foci that are elements of the class described by the unary predicate Φ, i.e., it is *wholly contained in the domain of* Φ:

Definition 39 *A sphere is contained in the domain of* Φ*:*
$ContainedSphere_{\Phi}^{t_1,t_2}(x, y) \leftrightarrow (x)InScope(t_1, t_2) \wedge (y)InScope(t_1, t_2) \wedge$
$(\forall y^*.(xy^* <_L xy \wedge (y^*)InScope(t_1, t_2)) \rightarrow \Phi(y^*))$

Similarly, a sphere may also be *self-connected* by a binary relation R:

Definition 40 *A sphere is self-connected by relation* R*:*
$ContainedSphere_{R}^{t_1,t_2}(x, y) \leftrightarrow (x)InScope(t_1, t_2) \wedge (y)InScope(t_1, t_2) \wedge$
$(\forall y^*.(xy^* <_L xy \wedge (y^*)InScope(t_1, t_2)) \rightarrow R(x, y^*))$

Note that since we require Φ (resp. R) to cover only the interior of the sphere, the sphere boundary does not have to be in the domain of Φ. This allows the sphere to "touch" the outside of Φ, and thus allows to formulate a finite variant of

an *open set* in point topology. Note also that each contained sphere has its own temporal scope and that x, y are moments inside this scope. The corresponding notion of a *closed contained sphere* includes the sphere boundary in the domain of Φ:

Definition 41 *A closed contained sphere:*
$cContainedSphere_{\Phi}^{t_1,t_2}(x, y) \leftrightarrow (x)InScope(t_1, t_2) \wedge (y)InScope(t_1, t_2) \wedge$
$(\forall y^*.(xy^* \leq_L xy \wedge (y^*)InScope(t_1, t_2)) \rightarrow \Phi(y^*))$

In order to capture the idea of "moving a sphere one step forward", we introduce an *overlap* relation between spheres. Two spheres (properly) overlap if they do not touch and are not disconnected. This means there is a focus u which is part of both spheres and does not lie on at least one sphere boundary, so that interiors must overlap. Note that this definition requires one of the two spheres to have a non-zero radius. This is because focus u is required to lie *inside* this sphere, and so this inside cannot be empty as in case of a degenerated sphere:

Definition 42 *Overlapping of two spheres in focus u:*
$(u)SphO(x, z, y, v) \leftrightarrow xu \leq_L xz \wedge yu \leq_L yv \wedge \neg(xu =_L xz \wedge yu =_L yv)$

In a similar way, we can define a temporal overlap relation between intervals:

Definition 43 *Temporal overlap of two intervals in focus v:*
$(v)tO(t_1, t_2, t_3, t_4) \leftrightarrow (v)InScope(t_1, t_2) \wedge (v)InScope(t_3, t_4)$

We "move a contained sphere one step forward" by requiring two contained spheres to overlap in a single focus with respect to the spatial as well as the temporal sense of "overlap" defined above:

Definition 44 *Moving a Φ-contained sphere one step forward:*
$CSStep_{\Phi}(x, y) \leftrightarrow \exists u, v, t_1, t_2, t_3, t_4.(v)tO(t_1, t_2, t_3, t_4) \wedge \exists z, w.$
$(u)SphO(x, z, y, w) \wedge ContainedSphere_{\Phi}^{t_1,t_2}(x, z) \wedge ContainedSphere_{\Phi}^{t_3,t_4}(y, w)$

Note that because of Definition 42, the two spheres of a step *can never touch but their interiors must overlap.* So their sum of foci must be strongly connected in the sense described above. Note also that x and y are moments that lie inside the respective temporal interval of the contained sphere. Overlapping temporal intervals therefore ensure that all moments in the interval from x until y and inside one of the spheres are also in the domain of Φ. Furthermore, if the step has a non-zero length, then one of the temporal intervals must also be non-zero, otherwise x and y would be equal as their temporal scopes need to overlap. Note also that this relation is *symmetric*, as we can easily exchange the two spheres centered on x and y with each other and immediately obtain $CSStep_{\Phi}(y, x)$.

We can now let a contained sphere "move" through a class, building a *chain of contained spheres.* Since our domain is finite, this movement is not continuous in a mathematical sense. It can be defined as the transitive chain of contained sphere steps concatenated in a finite sequence (compare Figure 6.6 and Definition 18):

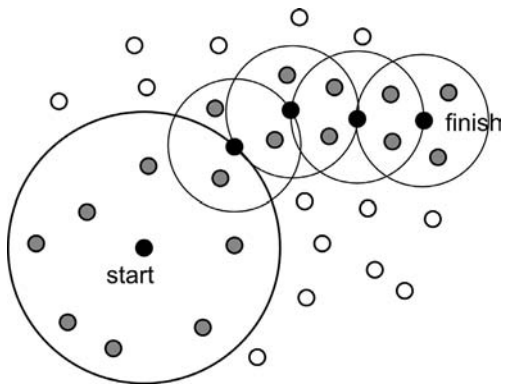


Figure 6.6: A chain of contained spheres generated by stepping through the foci of a class z (black and gray dots). The interior of every sphere of this chain contains only foci in z.

Definition 45 *Connected by a chain of contained spheres:*
$CSphereC_{\Phi}(x, y) \leftrightarrow CSStep_{\Phi}^{TC}(x, y)$

Note that this relation is *transitive and symmetric* by construction. It asserts that there is a series of overlapping contained spheres from x to y, and that the interior of each sphere is in the domain of Φ during a certain interval due to Definition 39. Note that we thus do not require the spherical neighborhoods inside this path to be always contained in Φ. A focus is contained in Φ only during the temporal scope of a sphere. This allows the sphere locations to be occupied by foci which are not Φ at other times, and thus the Φ-coverage to change in time and space.

Obviously, the same can be done with the closed variant of a contained sphere in Definition 41. In the following, we will denote closed variants of contained sphere-based definitions with the prescript c, for example $cCSStep_{\Phi}$ and $cCSphereC_{\Phi}$.

It is interesting to study the formal properties of these notions. I will prove now that a simple model of a contained sphere chain is just one non-degenerated sphere. *CSphereC* is not reflexive on foci of a single locus, since this would require degenerated spheres with zero radius, which is excluded by Definition 42. Therefore the following is also the minimal model. First, I prove that a non-degenerated contained sphere is connected to all foci that lie inside of it:

Theorem 17 *Self-connectedness of contained spheres:*
$ContainedSphere_{\Phi}^{t_1,t_2}(x, y) \wedge xz <_L xy \wedge (z)InScope(t_1, t_2) \rightarrow CSphereC_{\Phi}(x, z)$

Proof By the condition above, there is a contained sphere on x bounded by y. Since $xz <_L xy \wedge (z)InScope(t_1, t_2)$ implies $\Phi(z)$ by Definition 39, it follows that $ContainedSphere_{\Phi}^{t_1,t_2}(z, z)$. Since $xz <_L xy$ also implies $\neg xz =_L xy$ by Definition 26, Definition 42 is satisfied if we take $u = z$ and consider as the second sphere the degenerated case zz[7]. So we have $(z)SphO(x, y, z, z)$.

[7]This works since Definition 42 requires only one sphere to be non-degenerated.

These conditions satisfy Definition 44, and it follows that $CSStep_{\Phi}(x, z)$. By Definition 45, we have $CSphereC_{\Phi}(x, z)$. *q.e.d.*

If we take $x = z$, we immediately get the result:

Theorem 18 *Reflexivity on non-degenerated contained spheres:*
$ContainedSphere_{\Phi}^{t_1,t_2}(x, y) \wedge \neg xx =_L xy \rightarrow CSphereC_{\Phi}(x, x)$

PROOF Just apply Theorem 17 to the condition and take $x = z$. *q.e.d.*

A general definition of *strong connectedness* of a class z is then straightforward. It just means that any two foci in z are connected by contained spheres (Figure 6.7). A minimal strongly connected set consists of only one non-degenerated contained sphere.

Definition 46 *Strong Connectedness:*
$StronglyConnected(z) \leftrightarrow \forall x, y \in z.CSphereC_{\lambda x.x \in z}(x, y)$

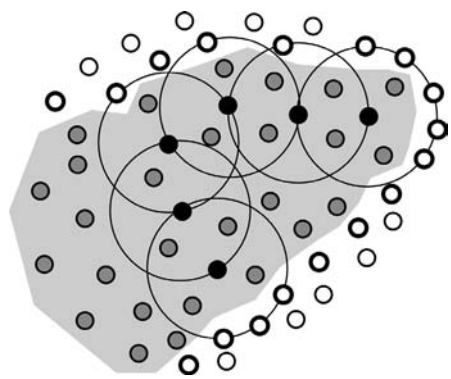


Figure 6.7: A strongly connected class z (gray and black dots) is self-connected by contained spheres. White dots in bold are the outerior boundary of z. These are touched by contained spheres.

Note that since contained spheres are 'open' sets (by Definition 39), they are allowed to touch a neighborhood of points in focus outside of z. Chains of contained spheres thus can indicate a *discrete boundary of* z. This idea will be used below to characterize visual surfaces.

Note that our constructed notion of strong connectedness remains incomplete compared to the standard notion in topology. It can only resemble it since it is not continuous. The degree of resemblance depends, on the one hand, on whether the observer has tested all foci of attention for lying inside one of these spheres. This means he has to construct for every point in focus a and every sphere xy at least one auxiliary point in focus y^* with $xa =_L xy^*$ on a line with xy that allows him to decide upon $xa <_L xy$. This reflects the idea that complex properties such as this one need to be explicitly established by an observer, otherwise he remains

unaware of them. On the other hand, resemblance also depends upon how many foci "samples" the observer has taken. This reflects the idea that the resolution of geometric and topological properties is a consequence of observation, as Frank [43] argued.

In order to mitigate this *effect of resolution* on topology, it is possible to introduce further axioms. In this case, one assumes there is always "enough" discrete attention available to observe a certain topological fact. For example, the following sentence about *touching contained spheres* would be an obvious theorem in continuous geometry. It says that a pair of touching (i.e. non-overlapping) spheres must be aligned with their touching focus. This alignment would be a provable fact in a continuous domain. In our discrete geometry, it needs to be added as axiom, in order to prevent spheres that overlap "unnoticed":

Axiom 24 *Touching spheres are always aligned with the touching focus:*
$ContainedSphere_{\Phi}^{t_1,t_2}(x^s, x) \wedge ContainedSphere_{\Sigma}^{t_3,t_4}(x^m, x) \wedge$
$(\neg\exists u.((u)SphO(x^s, x, x^m, x))) \rightarrow OnL(x^s, x, x^m)$

6.3.2. Media, substances and bodies as wholes under simple affordance

I assume that humans can directly perceive whether the environment affords a certain type of action, and are thereby able to *individuate bodies, media and their surfaces.*

Media. We can define media as unified wholes under simple affordance primitives. Humans can perceive a multitude of such simple affordances. One example are *locomotion affordances* for actions like diving, walking or driving.

Let us assume a primitive action type like *Diving*, standing for the locomotion of some body z in a fluid. The observation predicate $(z)DoDiving(x, y)$ expresses that the observer has detected the presence of a continuous diving action at foci x and y. $(z)AffordsDiving(x, y)$ asserts that the same kind of locomotion is present as a potential action. We may therefore define a *Fluid* as a medium, i.e., a maximal self-connected class of observed potential divings:

Definition 47 *Fluid medium:*
$Fluid(e) \leftrightarrow Whole^{F}_{AffordsDiving^*}(e)$

Where $AffordsDiving^*(x, y) \leftrightarrow \exists z.(z)AffordsDiving(x, y)$. This practically means that any foci of attention that are reachable via diving from this medium are part of it. Since affordance wholes are based on a partial equivalence relation by Axioms 9 and 10, media of a certain affordance type are equivalence classes and therefore *mutually exclusive*[8]. Note also that media defined in that way do not

[8]In contrast, media of different affordance type may overlap.

have any perceivable temporal boundaries if the affordance continues to exist. This is the case if the diving affordance can be successfully simulated during the whole of the attentional scope F. Therefore, I suggest to regard media as *endurants* from an ontological viewpoint [109].

Locomotion media. As discussed in the previous subsection, we perceive locomotions in terms of strongly connected paths of bodies observed for a continuous time interval. This impression of corporeal continuity exists despite the discreteness of attentional selection. It is enforced by *Gestalt completion* mechanisms, for example apparent motion (compare Section 3.2). Even if the moving body is *occluded for a limited time*, e.g., if it moves behind a wall or if it temporarily shifts out of attentional scope, we still have the impression of continuity. I propose therefore to model locomotion affordances as being strongly connected.

We can express this in the following schema. For reasons of generality, I suggest a meta-variable *Moving* ranging over all types of locomotion actions, such as *Diving*. We require that *the simulated action of a locomotion affordance is strongly connected.* For example, if we simulate a running, then we can imagine a still in which all foci on the runner's body are strongly connected, i.e., the body is connected by contained spheres. Since the movement is continuous, the next still shares some part in which contained spheres overlap, and so forth. The running therefore is *self-connected* with respect to *chains of affordance-connected spheres* (compare Definition 40):

Axiom Schema 11 *Afforded locomotions are strongly self-connected:*
$AffordsMoving^*(x, y) \rightarrow CSphereC_{AffordsMoving^*}(x, y)$

Locomotion media are therefore strongly connected:

Theorem 19 *Locomotion media are strongly connected:*
$Whole^F_{AffordsMoving^*}(e) \rightarrow StronglyConnected(e)$

Media of illumination and substances. According to Gibson, substances are not directly perceivable, only via the perception of a visual medium and its surface [50]. Just like walls are perceivable as obstacles to viewing. I suggest therefore to derive the notion of substance based on a *medium of illumination.* Alternatively, one might broaden the notion of substances based on other kinds of affordances. For example, the *Fluid* medium defined above has spatial boundaries at those locations where the diving stops. *I will restrict my understanding of surfaces and substances in this thesis to the visual domain*, but I expect that those shifted senses can also be very useful.

Beyond each visual surface, there lies a *substance.* Substances are the visual occluders of the environment. In order to describe this, we need another affordance predicate that allows to express whether some focus is visible, i.e., whether it affords being seen from some other point of view: $(z)AffordsSeeing(x, y)$ asserts

that it is possible see a point in focus z while looking from the point in focus at x and y. Note that, since this is an affordance assertion, it does not mean that the observer has actually taken the positions at x and y, but only that he successfully imagined taking them. More correctly, the positions x and y are eye positions (compare Figure 6.8) of an imagined observer.

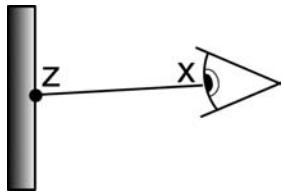


Figure 6.8: The observation primitive $(z)AffordsSeeing(x, x)$.

From a given point of view, an observer can see a surrounding configuration of non-occluded surfaces. Gibson called this the *optic array* (compare Figure 6.9). All foci in between the eye and a visible focus in this array are seen as well, and thus belong to a common medium of illumination. Inside an optical array, there is a transparent optic sphere surrounding the eye (compare Figure 6.9). An illuminated medium can be thought of as a configuration of such spheres, where each sphere stands for a transparent hull surrounding a potential viewpoint of an observer. In order to describe an illuminated medium, I propose to introduce a slightly more general notion of an *optic sphere*, i.e., a sphere that contains only foci of attention that are visible from somewhere during a certain time interval:

Definition 48 *Visible foci:*
$Visible(x) \leftrightarrow \exists z.(x)AffordsSeeing(z, z)$

Definition 49 *Optic sphere:*
$OpticSphere^{t_1,t_2}(x, y) \leftrightarrow cContainedSphere^{t_1,t_2}_{Visible}(x, y)$

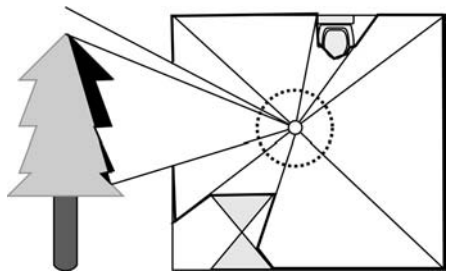


Figure 6.9: Gibson's *optic array* is a configuration of visual surfaces projected to a point of view in the environment [50]. Surfaces are not projected if they are occluded. I propose the notion of an *optic sphere*, in which every focus is visible, i.e., projected to the point of view.

An optic sphere is a sphere contained in an illuminated medium. By Definition 49, the smallest optic sphere is degenerated and consists of just one visible focus of attention. We construct an illuminated medium by realizing "visual steps" of overlapping optic spheres:

Definition 50 *Moving an optic sphere one step forward:*
$VisualStep(x, y) \leftrightarrow cCSStep_{Visible}(x, y)$

We can now define whether two foci are *connected by a chain of optic spheres*. This expresses that an observer can imagine to follow a path of overlapping optic spheres from one point of view to another one. Since this path is strongly connected by Definition 42, there is a *transparent corridor*, not only a mutually visible focus, which connects two successive spheres in this chain:

Definition 51 *Connected by a chain of optic spheres:*
$VisibilityC(x, y) \leftrightarrow cCSphereC_{Visible}(x, y)$

Note that all properties of contained sphere connectedness (compare Section 6.3.1) apply.

A *medium of illumination* can then be defined as a whole in which each pair of foci is connected by a chain of optic spheres:

Definition 52 *Medium of illumination:*
$IlluminatedM(z) \leftrightarrow Whole^{F}_{VisibilityC}(z)$

Such a medium of illumination can be explored by an observer in such a way that he follows observable positions of optic spheres, until he observes a surface where this chain must stop. For example, the explorable illuminated medium surrounding me shrinks to the interior of my flat if I close the outer door and the blinds. Note also that the smallest illuminated medium consists of a single non-degenerated optic sphere by Theorem 18. The illuminated medium can also change its position in time, as optic sphere locations may be occluded by substances at times outside the temporal scope of the sphere.

A substance is every nonempty class of foci which are invisible, i.e., which do not afford being seen from somewhere:

Definition 53 *Substances:*
$Substance(x) \leftrightarrow Class(x) \wedge (\forall y.y \in x \rightarrow \neg Visible(y))$

Humans can focus attention on substances even though they cannot perceive them. For example, it is possible to focus attention on an imagined point that lies two meters beneath a wall. It is clear that every focus must either be visible or not, and thus is part of a substance or illuminated medium, but not both.

Surfaces and bodies. Humans can recognize strongly connected substances bounded by a visible surface. These are called *bodies*. We can conceive bodies as *substance connected wholes*.

Definition 54 $Body(x) \leftrightarrow Whole^{F}_{CSphereC_{\neg Visible}(x)}$

Since *CSphereC* is transitive, its is provable that two bodies can never overlap, i.e., humans can take at most one body into focus at a time[9]:

Theorem 20 *Bodies never overlap:*
$Body(x) \wedge Body(y) \wedge \exists z.\,(z \in x \wedge z \in y) \rightarrow x = y$

PROOF By the condition and Axiom 7, $\forall a.a \in x \rightarrow CSphereC_{\neg Visible}(a, z)$ as well as $\forall b.b \in y \rightarrow CSphereC_{\neg Visible}(z, b)$. By transitivity of *CSphereC* due to Definition 45, also $CSphereC_{\neg Visible}(a, b)$. Since $\neg Visible(a)$ and $\neg Visible(b)$, by Definition 8, $a \in y$ and $b \in x$. By Axiom 1 it follows that $x = y$. *q.e.d.*

Surfaces must be identifiable by foci of attention, and therefore must have a *granular thickness*. This avoids the philosophical question whether an infinitely thin topological boundary belongs to a region or to its complement (see Casati and Varzi [26]). We take Borgo's view advocated in [18] and assume that every individuated body has its own surface, which is simply a thin layer of foci making up the visual boundary of the body or of the adjacent medium.

How do surfaces relate to substances and illuminated media? I have introduced the observation predicate *SC* for the individuation of a visual surface (Section 5.1.3). A *visual surface* can simply be conceived as a *surface connected whole*:

Definition 55 *Visual Surface:*
$Surface(x) \leftrightarrow Whole^{F}_{SC}(x)$

Above all, it is clear that every perceived surface affords being seen. Furthermore, surfaces are at the border of being seen, i.e. they border an illuminated medium. They also border a single strongly connected substance, of which they are said to be the surface[10]. We express this by requiring surface connected foci x, y to be touched by a chain of optic spheres as well as by a chain of substance contained spheres. This latter notion of touching is definable more generally (compare Figure 6.10):

Definition 56 *Touched by a contained sphere chain:*
$CSCTouched_{\Phi}(x, y, x^s, y^s) \leftrightarrow \exists t_1, t_2, t_3, t_4.CSphereC_{\Phi}(x^s, y^s) \wedge ContainedSphere^{t_1,t_2}_{\Phi}(x^s, x) \wedge \neg x^s x =_L x^s x^s \wedge ContainedSphere^{t_3,t_4}_{\Phi}(y^s, y) \wedge \neg y^s y =_L y^s y^s$

[9] An equivalent proof applies also to illuminated media.

[10] This fact might be used to consider surfaces as definable from illuminated media, as in [158]. I consider them rather as based on their own primitives, since I argued in Section 5.1.3 for a fundamental underlying Gestalt mechanism.

Note that the touched foci x, y are always inside the temporal scope of the adjacent sphere due to Definition 39, but lie on the boundary of the touching spheres on x^s and y^s. Note also that the touching spheres are *non-degenerated*. We will write *cCSCTouched* for the closed version of this topological relation.

The notion of surface connectedness with respect to its bounding substance and illuminated medium can now be specified by the following axiom:

Axiom 25 *Surfaces are the visible border of a substance and an illuminated medium:*
$SC(x, y) \rightarrow \exists x^s, y^s, x^m, y^m.CSCTouched_{\neg Visible}(x, y, x^s, y^s) \wedge$
$cCSCTouched_{Visible}(x, y, x^m, y^m)$

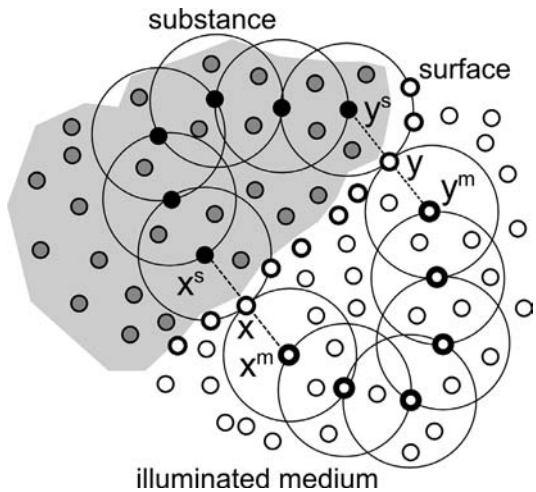


Figure 6.10: Surface connected foci x and y are connected by two touching sphere chains, one is open and inside a substance, the other is closed and consists of optic spheres.

We can now prove that surfaces are visible:

Theorem 21 *Surfaces are visible:*
$SC(x, y) \rightarrow Visible(x) \wedge Visible(y)$

Proof By Axiom 25, $cCSCTouched_{Visible}(x, y, x^m, y^m)$, thus $cContainedSphere^{t_1,t_2}_{Visible}(x^s, x)$ by the closed version of Definition 56. By Definition 41, $(x)InScope(t_1, t_2)$, and because trivially $x^s x \leq x^s x$ and the sphere is closed, it must be the case that $Visible(x)$. The same result is obtained for y by symmetry. *q.e.d.*

Note that although substance and surface connectedness are thus closely related, surface connection expresses in addition that foci are connected by a continuous visual corridor. This must not be the case for two substance connected surfaces: consider, for example, two foci on opposite surfaces of the wall of a closed room.

They are substance connected, but not surface connected, because they are not connected by an illuminated medium.

In order to prove some obvious properties about surfaces, we need to further explore the connection between surfaces, substances and illuminated media. For this purpose, we have to mitigate the resolution effect of discrete geometry again (compare Subsection 6.3.1). We need to exclude the case that an optic and a substance contained sphere can overlap[11].

Axiom 26 *Optic and substance spheres never overlap:*
$ContainedSphere^{t_1,t_2}_{\neg Visible}(x^s, x) \wedge OpticSphere^{t_3,t_4}(y^m, y)$
$\rightarrow \neg\exists u.(u)SphO(x^s, x, y^m, y)$

It is provable now that two substance (similarly, two optic) spheres touching in one focus on a surface must overlap:

Theorem 22 *Two substance (optic) spheres touching a surface in one focus must overlap:*
$SC(x, x) \wedge ContainedSphere^{t_1,t_2}_{\neg Visible}(x^s, x) \wedge ContainedSphere^{t_3,t_4}_{\neg Visible}(x^{ss}, x)$
$\rightarrow (x^s)SphO(x^s, x, x^{ss}, x) \vee (x^{ss})SphO(x^s, x, x^{ss}, x)$

Proof By Axiom 25, there is an x^m with $OpticSphere^{t_1,t_2}(x^m, x)$. By Axiom 26, the two substance spheres cannot overlap with this optic sphere, and thus by Axiom 24, they must be aligned with it, i.e. $OnL(x^s, x, x^m)$ and $OnL(x^{ss}, x, x^m)$. Also, x^m cannot coincide with x, i.e. $\neg x^m x =_L x^m x^m$, because the touching optic sphere must be non-degenerated by Definition 56. Thus, by Axiom 21, $OnL(x, x^s, x^{ss}) \vee OnL(x, x^{ss}, x^s)$. Thus, $x^{ss}x^s \leq_L x^{ss}x \vee x^s x^{ss} \leq_L x^s x$ by Definition 25. Furthermore, x coincides with neither x^{ss} nor x^s. For suppose the contrary, then by Definition 39 of a contained sphere and the condition of Theorem 22, x would be non-visible, but we know that x is visible by Theorem 21 and $SC(x, x)$. Therefore, $\neg x^{ss}x^s =_L x^{ss}x$. By Definition 42, therefore $(x^s)SphO(x^s, x, x^{ss}, x) \vee (x^{ss})SphO(x^s, x, x^{ss}, x)$. *q.e.d.*

The idea of a surface implies that it is *the surface of a unique body*. We can prove now that every surface is a surface of at most one body. Whether this body exists is a matter of reification in our constructive framework.

Definition 57 $SurfaceOf(s, b) \leftrightarrow Body(b) \wedge Surface(s) \wedge$
$\exists x, x^s, t_1, t_2.x^s \in b \wedge x \in s \wedge ContainedSphere^{t_1,t_2}_{\neg Visible}(x^s, x)$

Theorem 23 *Surfaces confine at most one body:*
$Surface(z) \rightarrow (\forall b, b^*.SurfaceOf(z, b) \wedge SurfaceOf(z, b^*) \rightarrow b = b^*)$

Proof By Definition 57 and the uniqueness condition, we have foci $x \in z$,$x^s \in b$ and $x^{*s} \in b^*$ with $ContainedSphere^{t_1,t_2}_{\neg Visible}(x^s, x)$ and $ContainedSphere^{t_3,t_4}_{\neg Visible}(x^{*s}, x)$. By Theorem 22, these spheres must overlap in x^s or in x^{*s}. Also, their temporal intervals overlap in x due to Definition 39. Therefore, by Definition 44, $CSStep_{\neg Visible}(x^s, x^{*s})$. By transitivity of $CSphereC$, $x^{*s}(x^s)$ must then be connected to all foci in $b(b^*)$, and thus $x^{*s} \in b \wedge x^s \in b^*$ by Definition 8 of a whole. By Theorem 20, we get the required result. *q.e.d.*

[11]Even though visibility and substances exclude each other, overlap may be possible if there happen to be no foci inside the substance sphere that are also on the optic sphere.

Similarly, it can be proved that illuminated media are unique on the opposite side of a surface.

And vice versa, the idea of a body implies a visual surface and thus an illuminated medium. Otherwise we would live in a dark world. We need to assert this by an additional axiom which excludes the possibility of a dark world:

Axiom 27 *Bodies have always visual surfaces:*
$Body(b) \rightarrow \exists s.SurfaceOf(s, b)$

The ground and the air. The different kinds of media often overlap and coincide for ecological reasons. This gives rise to more special kinds of media. For example, two central meaningful things of Gibson's environment can now be specified. The *air* may be defined as the intersection of an illuminated medium and a medium for breathing. The latter is based on a further affordance type *Breathing*, as part of a binary affordance predicate *AffordsBreathing*. It relates two foci of attention that are focused on an imagined breathing human being:

Definition 58 $AffordsLiving(x, y) \leftrightarrow VisibilityC(x, y) \wedge AffordsBreathing(x, y)$

Definition 59 $Air(a) \leftrightarrow Whole^{F}_{AffordsLiving}(a)$

Note that an air medium is strongly connected and maximal, but the illuminated medium may extend beyond it. For example, my room and the outside are visibility connected through my glass window, but glass is not a medium for breathing, and thus it splits the environment into two air media.

The ground is a *rigid body*, i.e. a substance humans cannot move through[12]. This may be observed in terms of a negated movement affordance predication[13]. Among the rigid bodies, it is a special one, namely the one under our feet. For simplicity reasons, I will just specify the notions of a *rigid body* and *rigid surface* based on rigidity connectedness:

Definition 60 $RigidityC(x, y) \leftrightarrow CSphereC_{\neg AffordsMoving^*}(x, y)$

Definition 61 $RigidB(b) \leftrightarrow Whole^{F}_{RigidityC}(b)$

Definition 62 $RigidSC(x, y) \leftrightarrow SC(x, y) \wedge$
$(\exists x^s, y^s.CSCTouched_{\neg AffordsMoving^*}(x, y, x^s, y^s))$

[12]There are of course movement types that allow to move through the ground. For example, it is a medium for moles.

[13]This is a preliminary solution. The absence of a movement affordance may not be enough since rigidity seems observable directly, not indirectly. Think about someone leaning against a wall.

6.4. Meaningful properties: waterdepth

It remains to show that the meaningful environment specified in Section 6.3 can be used to describe geodata. We begin by reconstructing a hydrological quality, as first suggested in [158].

Several characteristics can be used to describe the notion of river or lake, such as water depth or salinity. Dependent on the application area (e.g., navigation versus fishing, or diving versus swimming), particular ranges of water depth are categorized by introducing symbols like deep or shallow. To enable interoperability between different information communities, we have to ground such symbols in perceptual operations.

How could a grounded semantic description for "water depth of a lake" look like? The identification of this particular depth is dependent on its host. Probst and Espeter [132] have given an answer to this question based on the host's dimensionality. The one-dimensional water depth quality of a lake at a certain region of its water body is considered to have a value, a *quale* in DOLCE's use of the term [47]. A water body has more than one such quale. But for the host as well as the quality, we need observable individuation criteria. We can then show how these qualities can be derived from observation predicates.

Furthermore, the necessary distinctions are observer as well as actor relative: for example, water is a medium for fish, whereas it is, by and large, a substance for us. A human's perception of deep or shallow depends on planned activities. The diver may call a lake's water depth of ten meters "shallow", whereas the novice swimmer would call two meters "deep".

To create a reproducible classification of depth, we construct these categories in terms of surfaces, attentional steps and affordances.

Waterbodies and watersurfaces. Consider the geographic category water body, which encompasses seas, lakes, rivers, ponds. Such an accumulation of water is a fluid, a medium for diving, but also a part of a visual substance for humans, i.e., a "body" in our sense. It has a non- or semi-transparent visual surface. When we look at a lake from above, it is embedded into the opaque body of the earth landscape and constitutes part of its surface. This may perhaps reflect the etymological roots of the word "waterbody". We therefore say that the waterbody "touches" this surface, and that the surface is the larger landscape body surface which "adjoins" the air (compare Figure 6.11)[14]. Furthermore, the landscape body includes the ground as well as the water body.

Definition 63 $(b)\,Touches(s)in(z) \leftrightarrow$
$z \in s \wedge \exists u, t_1, t_2 . u \in b \wedge ContainedSphere^{t_1,t_2}_{\lambda x.x \in b}(u, z)$

[14]This simply means that there is some open contained sphere in the waterbody touching the surface and similarly, a closed contained sphere in the air touching every locus on the surface.

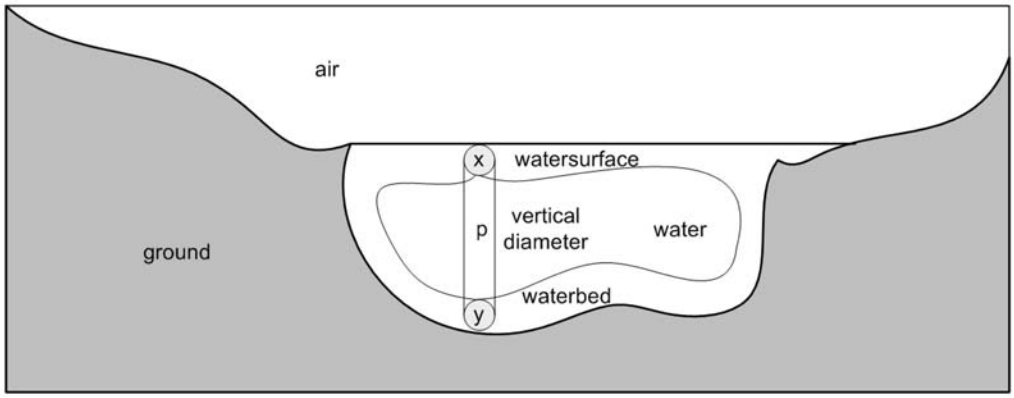


Figure 6.11: Water depth is the length of a vertical diameter path in a water body.

Definition 64 $(s)cAdjoinsTo(a) \leftrightarrow$
$(\forall x.(x \in s) \rightarrow \exists y, t_1, t_2.(y \in a \wedge cContainedSphere^{t_1,t_2}_{\lambda x.x \in a}(y, x)))$

Definition 65 *Water body b embedded in landscape ls with surface s:*
$Waterbody(b, s, ls) \leftrightarrow Fluid(b) \wedge \exists g.RigidB(g) \wedge Body(ls) \wedge g \subset ls \wedge$
$b \subset ls \wedge Surface(s) \wedge SurfaceOf(s, ls) \wedge \exists z.(b)Touches(s)in(z) \wedge$
$\exists a.Air(a) \wedge (s)cAdjoinsTo(a)$

A water surface is that part of the landscape surface which is touched by a water body:

Definition 66 $WaterSurfaceOf(s^*, b) \leftrightarrow$
$\exists s, ls.Waterbody(b, s, ls) \wedge (\forall z.(b)Touches(s)in(z) \leftrightarrow z \in s^*)$

Waterdepth of a waterbody. Informally, in the meta-data of a database, we could say that water depth of a river is the vertical distance measured between a locus on the water surface and the river bed. Once a discrete length space for steps is established as in Section 6.2, we can define these notions from the meaningful environment (compare [158]).

Let us start by defining a diameter of some thing in the meaningful environment: A *diameter* of class r is any maximally extended straight segment *which is contained in* r:

Definition 67 *Diameter:*
$Diameter(x, y, r) \leftrightarrow (\forall z.OnL(x, z, y) \leftrightarrow (z \in r \wedge (z \cong_y x \vee z \cong_x y)))$

We define a *depth of a thing* as the length of a diameter which is vertically aligned[15]:

[15]This is a simplification. The vertical calibration of such a line of depth measurement, e.g.

Definition 68 *k is a depth of r at positions x, y:*
$Depth(k, x, y, r) \leftrightarrow Diameter(x, y, r) \wedge distance(x, y) = k \wedge VertAln(x, y)$

We can now state that a water depth is a depth of a water body touching the water surface. There is a potentially infinite number of water depths for a water body (compare [132]):

Definition 69 *k is a water depth of r at positions x, y:*
$Waterdepth(k, x, y, r) \leftrightarrow Depth(k, x, y, r) \wedge \exists s, v. WaterSurfaceOf(s, r) \wedge$
$v \in s \wedge OnL(y, x, v)$

Sticks and sonars both can be considered devices that produce realizations of such a virtual water depth path. They allow to construct diameters because they can be extended until they reach the ground, i.e., until they stop moving. Note that in our discrete domain of observations, this depth varies with the resolution, i.e. with the closeness of foci of attention x approximating the surface s. Note that in common hydrological data models, waterdepth or waterlevel are attributes of some relative river location. The proposed definitions are comparably complex, but simply because they require to be explicit about the different types of observation involved, including watersurface, ground, air and water depth path.

6.5. Meaningful media: roads and places

Different kinds of media ground different geographic categories. In Chapter 7, I will show how a locomotion medium can be used to ground road network data. A more general *support medium* is sketched in the next paragraph. I will then roughly outline the idea to conceive of a place in terms of a medium. Both concepts are central to geographic information.

Supported locomotion. Predicating affordances in the meaningful environment should be interpreted as an observable fact. As Gibson argued, this fact is hardly reducible to other facts, and therefore may not be definable. However, it is observable by humans in the sense that it is an inter-subjectively available result of their sensori-motor simulation, and therefore it is at the roots of our method for grounding databases.

Roads are configurations of surfaces, air and substance parts that afford *supported locomotion*: cars, bicycles and pedestrians do not move through the concrete bodies of roads, nor do they fly arbitrarily around in the air, but they move through a piece of air right above a supporting surface such that they are constantly in touch with it [83].

---

a measuring stick, is normally done above the water level. More correct geodetic leveling even requires the line to be a non-straight path that approximates curved gravity lines, compare the paragraph about calibration in Section 5.1.5. This last solution was used in [158], but is omitted here for convenience.

We may therefore use more specific affordance primitives based on *AffordsLiving* (underlying air media, see p. 6.3.2) and *AffordsMoving* (see p. 6.3.2) for road network representations. The approach is well suited to capture the multimodality of traffic, since all modes of traffic, e.g. walking, driving, diving and flying, have such observable affordances and corresponding media.

For *supported locomotion* [153], we assume that there is a subtype *SupportL* of *Moving* for it. The affordance predicate $(z)AffordsSupportL(x, y)$ means that *foci* x, y *are connected by a simulated locomotion which is supported by a flat surface located at* z. I omit any formal description of the geometric support surface property, because it is not necessary as a premise for the subsequent arguments. Two aspects of its intended meaning can however be easily expressed by asserting that x, y are part of an air medium and that the *support* is a *rigid reference surface underneath* (compare Definition 62):

Axiom 28 *Locomotion supported by a surface in z:*
$(z)AffordsSupportL(x, y) \rightarrow AffordsLiving(x, y) \wedge \exists u.RigidSC(z, u) \wedge$
$VertAln(x, z) \wedge VertAln(y, u)$

Note that since *SupportL* is a movement type, Axiom Schema 11 applies to it. Therefore, supported movement affordance tuples are connected by a self-connected path, and so their media are strongly connected. A *support medium*, a medium for supported locomotion, extends through an air medium as long as there is a rigid surface underneath:

Definition 70 *Support media:*
$SupportM(x) \leftrightarrow Whole^{F}_{AffordsSupportL^*}(x)$

Meaningful places. While computational approaches to place, such as OSM Points of interest or gazetteers, use simplistic data models, the important notion of place as a semantic referent [70] encompasses a large variety of meanings, ranging from spatially referenced locations, e.g., the Royal Observatory in Greenwich, through objects such as vessels, to point-like experiential spaces, and social handles such as the notion of home [29]. The main hypothesis, proposed already by [75] and advocated by myself and others in [160] and [152], is that many of these crucial aspects of place can also be captured by affordances. However, this is work in progress, and so my discussion will be informal.

Being able to distinguish places according to their affordance is of immediate use for describing Points of Interest (POI) more appropriately. For example, many cafés in Europe are open late and also serve alcohol, so that they would be better described as bars in the evening. Annotating such POI with either `amenity=cafe` or `amenity=bar`, which is the current practice in OSM [160], therefore only tells half the truth and hides useful information. Post offices offer a number of different services in Germany, including banking facilities or the

possibility to buy stationary. Describing such a POI only as `amenity=post_office` does not give credit to all these different functionalities and makes it hard for other users to figure out what kind of services they can expect at this place [160].

To overcome these problems, I proposed in [152] that a place may be individuated in terms of some special kind of medium. Places satisfy all characteristics

Figure 6.12: A marketplace is a medium that moves with its vendors (by kind permission of Julianna Kuruhiira and Phil Bartle ©, cf. `http://www.scn.org/cmp/`).

of media (compare [29]): They can be traversed, but can also be located, and can move themselves if the perceived affordance that constitutes them appears or disappears at some location. For example, the sea as a place for organisms can move across river banks. And the place of Horatio Nelson's death during the battle of Trafalgar is the H.M.S. Victory, which is currently located at Portsmouth, England [70]. Furthermore, media are primary categories of perception, just as places, and as such can be also considered subjects of human affection and identification. Finally, media may be defined in relation to some reference surface which is interwoven with their primary affordance [152].

Take the example of a market place (Figure 6.12), which may not be delimited by any wall or sign. Markets are media that primarily afford people to move in communication distance to vendors. In this case, the identifiable surface is made of a configuration of other human bodies and the perceivable relation is one of meeting distance, while the location of the market place is the maximal sum of all locations one can move through in communication distance to those vendors.

6.6. The meaningful environment revisited

In this chapter, I have illustrated how the observation predicates introduced in Chapter 5 can be used to reconstruct essential categories of our observable human environment. The challenge consists of two main strands. First, our domains of interpretation F, which is a memorized window of attention, and D, which also includes constructions, are finite, but most mathematical abstractions employed to describe the environment are infinite. And second, we cannot presume absolute but need to build relative experiential geometries. In effect, geometrical

measurement needs to be conceived as a construction that is logically *subordinate* to affordance and surface perception. And analogously for the major categories of Gibson's meaningful environment, such as media, substances and surfaces.

Name	Formal symbol	Definition
length order	$\leq_L$	25
length addition	$+$	27
congruence of angles	$=_A$	30
angle order	$\leq_A$	31
length	*Length*	33
contained sphere	$ContainedSphere_{\Phi}^{t_1,t_2}$	39
contained sphere connectedness	$CSphereC_{\Phi}$	45
strong connectedness	*StronglyConnected*	46
fluid medium	*Fluid*	47
visibility	*Visible*	48
optic sphere	$OpticSphere^{t_1,t_2}$	49
medium of illumination	*IlluminatedM*	52
substance	*Substance*	53
body	*Body*	54
surface	*Surface*	55
air	*Air*	59
rigid body (ground)	*RigidB*	61
waterbody	*Waterbody*	65
watersurface	*WaterSurfaceOf*	66
diameter	*Diameter*	67
waterdepth	*Waterdepth*	69
support medium	*SupportM*	70

Table 6.1: Inventory of meaningful things.

Table 6.1 is an *inventory* of meaningful things that were reconstructed in this chapter. I also demonstrated that topological and geometric properties, such as strong connectedness of media and bodies, can be understood in a finitist manner. However, there is a price to pay for finitism. Major theorems of geometry, such as metric properties of length measurement, can only be proved based on *auxiliary constructions*. And we have seen that topological inferences *depend on resolution*, which required us to add axioms for compensation.

Even though the notions of continuity and density seem intuitively understandable because they originate in perception, substantial constructive effort is required to make them available on an explicit level. No doubt, they are useful fictions for spatial reasoning. However, reconstructing these mathematical notions requires extra effort. Their reconstruction could be based on the idea of potentiality [61], or on analogical reification and recursive operational representations (as suggested in Sections 4.1.2 and 4.2). This is considered future work.

The illustrated way of reconstructing our environment is more complicated than ordinary ontological or mathematical approaches. But it has the advantage of exposing the roots of reference of our language, which are usually blanked out.

Chapter 7

A grounded theory of road networks

Alles was überhaupt gedacht werden kann, kann klar gedacht werden.
Alles was sich aussprechen läßt, läßt sich klar aussprechen.

— L. Wittgenstein [206]

Road network data is among the most widely known types of geodata, due to its widespread use in navigation and orientation applications. The demand for business applications based on road networks largely exceeds their original application in navigation systems [111]. Non-standard applications require non-standard (non-available) labels and attributes. User generated geodata, like Open Street Map, has become a competitor. Here, labels can be freely attached to geometries without any top down constraints. Yet, the many quality issues[1] make semantic heterogeneity of such grass root labeling a big challenge.

Even though informal standards for road network labeling are available (e.g. the GDF standard, ISO 14825:2004), geometries and category labels in different commercial road network databases are far from denoting equivalent things. Furthermore, the available categories seem to be chosen haphazardly and often lack high-level features, like junctions and roads, as well as labels for complex generalization tasks [157]. Data models are usually *incomplete*: Essential properties are not expressed as facts in the data [153, 129]. One would prefer a method for grounding road network categories in *observation procedures* in order to support more general data mappings. If road network categories were definable from directly observable primitives, this would provide a common approach to semantically annotate and compare the contents of heterogeneous road network databases and services.

Observable primitives for the individuation of road networks and their features can be found in Gibson's *meaningful environment*. I account for road networks using an affordance primitive for *turn-compliant locomotion*.

[1] `http://wiki.openstreetmap.org/wiki/Category:Quality_Assurance`

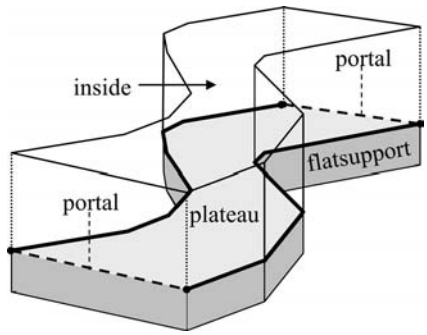


Figure 7.1: Illustration of a channel. It is part of an air medium which is located just above a flat smooth surface. Furthermore, it restricts supported movements to enter and leave at separate portals by convention.

In this chapter, I will demonstrate how *channel networks*, that were first proposed in [153] and [154], can be constructed from this observation predicate (Section 7.1). I will also show how road network data models can be interpreted into this theory of channels (Section 7.2). The theory can be used to define road network features like junctions (Section 7.3). Last but not least, based on the work in [156], I will apply the junction definition to examples and shortly illustrate its use by a junction search tool for OSM that was built on it (Section 7.4).

The presented theory of channel networks extends the work in [154] by *defining* the concept of a channel, while the definition of a junction and its application to OSM is a short version of the work already published in [154, 129, 156].

7.1. Channels and what they afford

Roads are parts of a medium of supported locomotion (Section 6.5). But roads additionally exhibit conventional *turn restrictions*, observable, e.g., by road signs or dotted vs. full lines on the side of a road surface. Turn restrictions are the main instrument for traffic planning, because they restrict the supported movements to a regulated more or less homogeneous flow into discrete directions. This is a case of *physically manifested social affordances*, because it presupposes the interpretation of symbols on the road surface.

I account for this complex affordance by a further movement type, called *TurnOffC*. It denotes a supported motion which is *in compliance with any turn restriction* (TOC movement):

Axiom 29 *Turn-compliant (TOC) movement affordance:*
$AffordsTurnOffC(x, y) \rightarrow AffordsSupportL^*(x, y)$

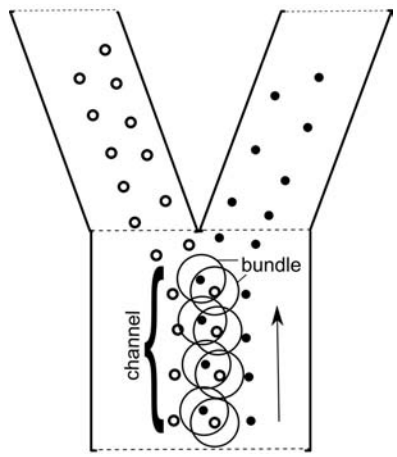


Figure 7.2: Channels can be observed as bundled sections of turn-compliant movement affordances. The figure shows two afforded movements, one depicted by filled dots, the other one depicted by empty dots.

I define a *channel* as a part of a TOC locomotion medium. The idea is that channels are compact spaces which force their afforded locomotions to move *into only one direction*, and to leave into or enter *from another channel* at portals (see Figure 7.1). How can this notion be constructed? I propose that observing a channel just means to identify a maximal bundle of overlapping TOC movement affordances, and to follow it until the bundle splits, i.e., until some movements are allowed to turn off. A particular channel provides space for one bundle of movements and ends at any turn off possibility.

I first define a *binary movement bundle* as a couple of spatially overlapping movement contained spheres[2], where each center is focused on a different simulated movement (compare Figure 7.2):

Definition 71 *Binary bundle of movements:*
$Bundle_{TurnOffC}(x, y) \leftrightarrow \neg AffordsTurnOffC(x, y) \wedge \exists z, w, u, t_1, t_2, t_3, t_4.$
$ContainedSphere^{t_1,t_2}_{AffordsTurnOffC}(x, z) \wedge ContainedSphere^{t_3,t_4}_{AffordsTurnOffC}(y, w) \wedge$
$(u)SphO(x, z, y, w)$

A bundle of movements is connected by a *transitive chain* of such binary bundles (compare Definition 18). This corresponds to a series of slightly shifted spatially overlapping movements:

Definition 72 *Bundle connectedness:*
$BundleC_{TurnOffC}(x, y) \leftrightarrow Bundle^{TC}_{TurnOffC}(x, y)$

[2]This means that they spatially overlap in a movement focus u which is not on both affordances. Otherwise, affordances would overlap themselves, which is excluded by Axiom 10.

The chain ends if the observer cannot successfully simulate another TOC movement anymore which spatially overlaps with the last one. A bundle is a maximal set of bundle-connected foci:

Definition 73 *Bundle:*
$Bundle_{TurnOffC}(b) \leftrightarrow Whole^{F}_{BundleC_{TurnOffC}}(b)$

Observing turn restrictions requires observing the possibility of *moving into a certain direction*. The directnedness of an affordance was not considered so far. Note that it is not implied by the order of foci of the affordance relation, because affordance assertions are symmetric. But it can be appropriately defined as *the direction into which the underlying locomotion was successfully simulated*, based on the *temporal order* of foci of attention that followed this simulated movement:

Definition 74 *Directedness of an affordance:*
$Directed_{Moving}(x, y) \leftrightarrow AffordsMoving^{*}(x, y) \wedge x \leq_{T} y$

Note that $\leq_T$ is antisymmetric by Axiom 4, so $Directed_{Moving}$ is not symmetric and can therefore express a direction.

If a movement bundle consists of precisely those continued movements that started from another bundle, we say it follows inside the same channel, and the two bundles are *channel connected*:

Definition 75 *Channel connectedness:*
$ChannelC(a, b) \leftrightarrow Bundle_{TurnOffC}(a) \wedge Bundle_{TurnOffC}(b) \wedge$
$(\forall x, y.(x \in a \wedge y \in b) \rightarrow (\exists z, u.(z \in b) \wedge (u \in a) \wedge$
$Directed_{TurnOffC}(x, z) \wedge Directed_{TurnOffC}(u, y)))$

Channels are maximal sums of foci connected in this way:

Definition 76 *Channel:*
$Channel(c) \leftrightarrow Whole^{F}_{\lambda x,y.(\exists a,b.(ChannelC(a,b) \vee ChannelC(b,a)) \wedge x \in a \wedge y \in b)}(c)$

Figure 7.2 illustrates this definition. Spatial compactness is assured since foci are connected by movement contained spheres (Axiom Schema 11). After some turn off possibility, there are movement opportunities that do not lie inside the same bundle anymore, because they are separated by some crutch. The foci beyond this point cannot be part of the same channel, and so the channel ends at this point. Note that since *AffordsTurnOffC* is transitive, channels cannot overlap in space and time.

We can now define a locomotion affordance relation between channels. A channel *leads to* another one if the first is connected to the second by a turn-compliant movement:

Definition 77 *LeadsTo:*
$LeadsTo(a, b) \leftrightarrow Channel(a) \wedge Channel(b) \wedge (a \neq b) \wedge$
$(\exists x, y.(x \in a) \wedge (y \in b) \wedge Directed_{TurnOffC}(x, y))$

This is the central notion that allows us to ground road network databases, since it captures the fact that one is allowed to turn off from one channel into another. As argued above, channels cannot overlap, and are *discrete*. Furthermore, *LeadsTo* is *irreflexive*, because channels always *lead to somewhere else*. I will illustrate these constructed notions in the following by comparing them to data models.

7.2. Affordance-based interpretation of road networks

In this section, I will illustrate how common road network data models can be interpreted into a channel network, which is constructible from channels. The presentation is a slight modification and synopsis of ideas first presented in [154].

7.2.1. Road network data models

One can think of channels as one side of a bidirectional road or a one-way street between two intersections. A channel is a part of a support-connected medium in which movements are restricted to only one direction *by convention*. This is the reason why in road network data models, channels are usually embedded into the plane by line segments whose end nodes indicate some kind of intersection. But obviously the data model needs to be more complex than an ordinary

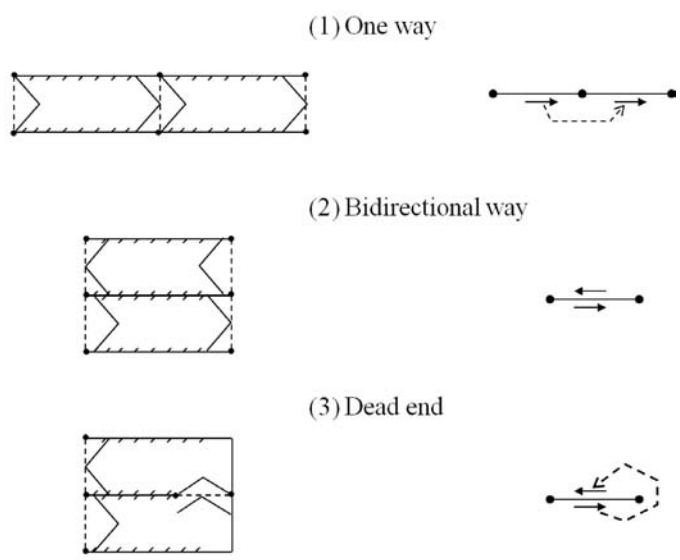


Figure 7.3: Channel configurations forming road network data elements.

graph: In Figure 7.3 and Figure 7.4, I have depicted example channel configurations (left) together with their equivalent *road network data models* (right),

denoting e.g. *one-way streets* and *dead ends* [157]. In the data model sub figures, the *LeadsTo* relation is depicted by dotted arrows, channels are represented as full arrows, and the embedded street segments are represented as a separate undirected graph with lines and nodes. Note how the road network data model

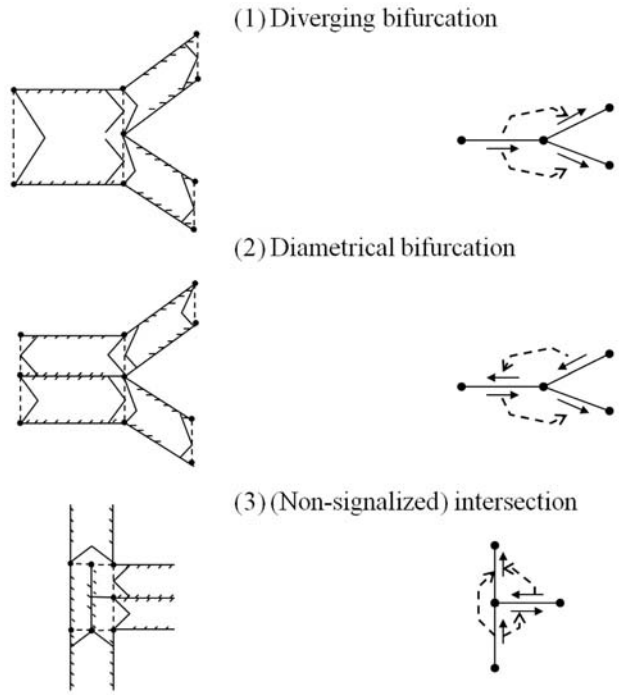


Figure 7.4: Channel configurations forming road network data elements.

(right) is supposed to be interpreted in the channel network theory (left): Any adjacent and consecutive pair of sides of embedded street segments which is not affected by turn prohibition is equivalent to a tuple of channels of the *LeadsTo* relation[3] [153]. The existence of turn restrictions has formal consequences: The embedded street segment graph (depicted by the undirected graph on the right hand side of Figure 7.3 and Figure 7.4) needs to be supplemented by a second relation on top of it (depicted separately as dotted arrow). This is because the LeadsTo relation would not be captured otherwise: the two branching edges of a *diametrical bifurcation* (Figure 7.4.2), for example, are connected at a common node, but do not lead to each other. Especially interesting are pairs of channels that are support connected *but neither one leads to the other*. A plausible model for this are neighboring channels built on the same surface, like the neighboring

[3]At signalized intersections, this includes a *temporary* channel. Temporary channels exist only during green phases and do not have data equivalents.

sides of a bidirectional road without a median strip (see the bidirectional way in Figure 7.3.2 or the diametrical bifurcation in Figure 7.4.2).

Hence, a road network can be appropriately represented by a subgraph of the *line graph*[4] of a street segment graph, i.e., a graph with nodes for channels and edges for those channel pairs where one channel leads to the other. Such a line graph offers also an appropriate way of modeling turning costs in route planning [205].

7.2.2. Channel networks and channel digraphs

Channel networks are media that afford *mutual reachability* by turn-compliant movements between its channels. The ternary relation
ReachableFrom is definable as the *transitive closure* of *LeadsTo* (compare Definition 18):

Definition 78 $(y)ReachableFrom(x) \leftrightarrow LeadsTo^{TC}(x, y)$

Note that the relation *ReachableFrom* is neither symmetric nor reflexive but transitive[5].

We can now define a *channel network* as a whole of channels with respect to reachability:

Definition 79 $ChannelNetwork(x) \leftrightarrow Whole^{Channel}_{ReachableFrom}(x)$

Note that *wholeness* requires for a channel network that reachability amongst its channels be *reflexive* (all channels of the network can be reached by navigation from themselves in the network) and *symmetric* (it is always possible to return to the origin). It also requires that the network is a class of channels, i.e., a second order class of classes. Furthermore, as was discussed in [153] in greater detail, every channel network must consist of at least two mutually reachable channels (minimal model): There must be at least one channel by definition of a whole (Definition 8), but because this channel has to be reachable from itself inside the network (Definition 79), and leadsTo is irreflexive by Definition 77, there has to be at least another non-overlapping one wich leads back to the first. A channel network also does not contain any *graveyards and factories*, channels without the possibility to leave or enter (compare [157]). This follows directly from Definition 7 and the minimal model.

The theory of channel networks satisfies a graph theory. I call this theory the *channel digraph*. Channel digraphs will be used in the following text and figures

[4]A line graph L(G) of a graph G is the graph on G-edges in which two G-edges are adjacent as vertices if and only if they are adjacent in G [31].

[5]There is an even stronger notion of reachability, which was called *2-reachability* in [154]. For convenience, I restrict this presentation to this simpler notion which captures most of the intention.

in order to describe our formalism on another abstraction level, and in order to make it more amenable to data structures and algorithmic standards that are based on graphs. Note that actually I do not leave the grounded FOL theory.

If we take the *set of channels* of a channel network as the set of vertices V, and the relation *LeadsTo* as the arc relation A (directed edges), then the resulting graph $D(V, A)$ is a directed graph without loops (arcs connecting one vertex to itself, due to Definition 77) and multiple arcs (arcs with the same incident vertices). For an arc $e = (x, y)$, I call the incident vertex x the initial, and y the terminal vertex of e[6].

7.3. Affordance-based definition of a junction

The theory of channel networks can be used to check and categorize road network data with respect to what their *observable* referents afford. In this section, I will illustrate categorization of the most prominent road network features, namely junctions[7].

7.3.1. Channel network features as induced subgraphs

An *induced subgraph* $S(U, O)$ of a channel digraph has a subset of vertices $U \subseteq V$ of the channel digraph, and the set of arcs with elements $e \in A$ connecting any pair $\{x1, x2\} \subset U$, is exactly its set of arcs, $e \in O$. A channel network feature, like a road or junction, is an induced subgraph of D, because the selection of channels should preserve the affordance relation between them.

The essential *affordance properties* of such induced subgraphs now seem to arise from the connection properties of some of their vertices, called *entries, exits and gates* (Figure 7.5). Consider all arcs from $A \backslash O$ that are incident with vertices in U (connecting S with its complement in D), and call them *in-/out bridges of* S. I call a vertex $v \in U$ an *entry* of S, if and only if it is a terminal vertex of an in-bridge, and an *exit* if and only if it is an initial vertex of an out-bridge. The set of bridge-incident vertices that are either entries or exits (but not both), is called *gates of* S. Gates allow traffic to either enter or leave the subgraph, but not both. Note that a gate forces a moving object traversing it to move on a walk of length ≥ 1 inside the subgraph S, because it cannot leave S immediately. In the following definition, the part-of relations were introduced in Definition 5, and x denotes a subgraph of a channel digraph.

Definition 80 $Entry(en, x, y) \leftrightarrow ChannelNetwork(y) \wedge PP(x, y) \wedge$
$en \in x \wedge (\exists z. \neg z \in x \wedge z \in y \wedge LeadsTo(z, en))$

Definition 81 $Exit(ex, x, y) \leftrightarrow ChannelNetwork(y) \wedge PP(x, y) \wedge$
$ex \in x \wedge (\exists z. \neg z \in x \wedge z \in y \wedge LeadsTo(ex, z))$

[6]For details see the presentation in [154].
[7]This is an abridged version of [154].

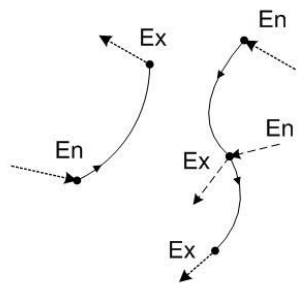


Figure 7.5: An induced subgraph of a channel network and its entries and exit vertices (indicated by dotted arrows, gates are narrowly dotted).

Definition 82 $Gate(e, x, y) \leftrightarrow (Entry(e, x, y) \wedge \neg Exit(e, x, y)) \vee (Exit(e, x, y) \wedge \neg Entry(e, x, y))$

7.3.2. Affordance based definition of n-way junctions

Now we are in a position to define a junction based on the actions it affords as part of a channel network. We can do this because our graph theory is grounded in locomotion affordances. An *n-way junction* is an *induced subgraph* of a *channel digraph*, which

1. affords $n - 1$ (n $\geq$ 3) navigational choices for n entries, as there is a walk from each of the n entries to each of n exits except into the opposite direction (*total reachability*),
2. affords movements to enter and leave through distinct entry and exit channels (*discreteness of navigational action*),
3. does not contain a smaller n-way junction (*minimality I*) and
4. has entries and exits with a *minimal vertex degree of two*. Entries have either more than one internal successor or an internal predecessor. Analogously, exits have either more than one internal predecessor or an internal successor (*minimality II*).

I discuss and motivate these properties in the rest of this subsection by analyzing a *median u-turn junction* (see Figure 7.6). At a median u-turn intersection, the main road, a dual carriageway, intersects the minor road, which is a bidirectional road. It is prohibited to turn off to the left from the major road or onto the major road at the intersection point. These left turns are only possible via u-turn channels (see data model in Figure 7.6). Channels are indicated by arrows alongside their embedded street segment lines. Note that the bidirectional street segments have two channels, one for each direction, while the dual carriageway consists of one-way street segments. The *LeadsTo* relation among channels is indicated by

gray curved arrows that lead from a channel to its successor. *Exits and entries* of the subgraph are numbered anticlockwise. Feature external channels are in gray. I will refer to the correspondent *channel digraph* in Figure 7.7 a, in which channels are depicted as vertices and the LeadsTo relation as arcs. The exits and entries in these figures have equivalent numbers.

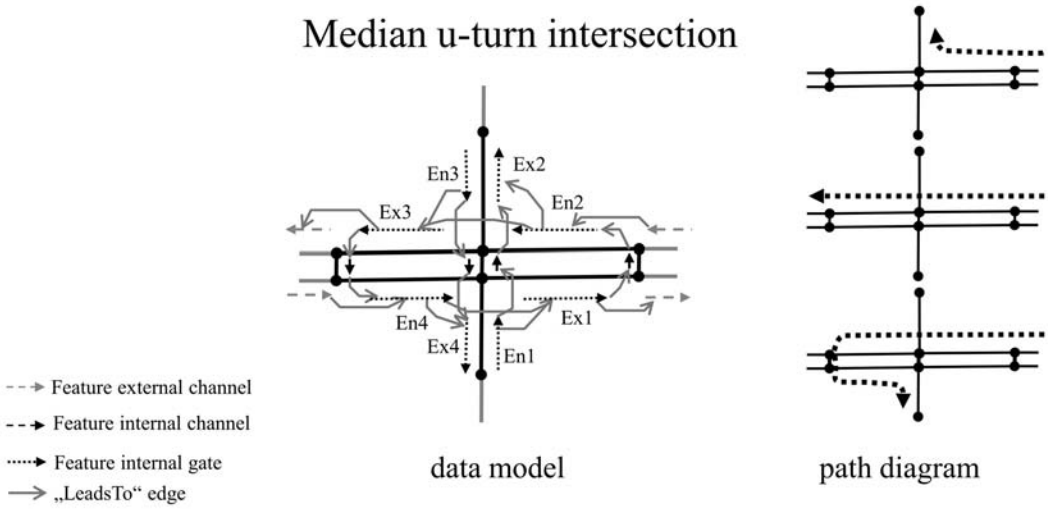


Figure 7.6: A median u-turn (Michigan left) intersection data model.

The first and most obvious property is that junctions afford paths from each entry to each exit (except the one in the opposite direction) (*total reachability*). Thus they enforce a navigational choice. If a driver has entered a junction, he or she is afterwards forced to take a directional decision by taking one of a set of $n - 1$ paths inside of it. Note that this property also allows drivers coming from different directions to take the same exit. Each entry is used by people driving into different directions, and each exit is used by people coming from different directions. Together with the minimality assumption, this property also implies that a junction is a *connected* subgraph. If one leaves out the dotted arc in Figure 7.7 b, then entries one, two and four are affected and lose their paths to exits three and four, so that this subgraph is not a junction anymore.

The second property, *discreteness of navigational action*, seems to be common to every *road network feature*, thus also to a road. If people are speaking of 'staying on or leaving a road/junction', I suggest that they mean *discrete actions*: The process of entering, staying on, or leaving a road/junction is unambiguous for a moving object inside of a channel. So roads and junctions are assumed to have *mutually exclusive entry and exit channels (gates)*. An obvious reason for this is that navigational actions are kept simpler and more transparent for other road users. The subgraph in Figure 7.7 c (in black) is a part of the median u-

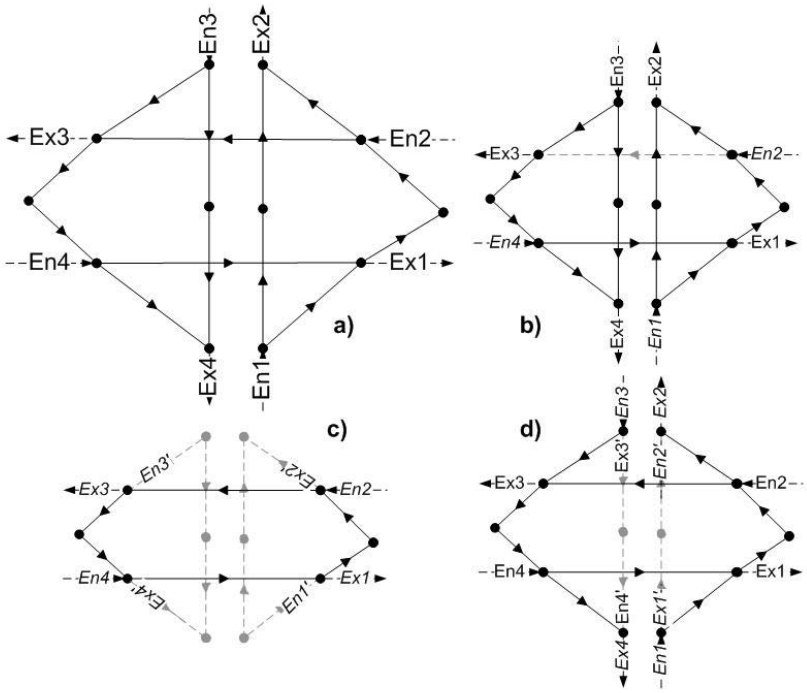


Figure 7.7: Illustration of junction properties by a *median u-turn* channel subgraph a). If one leaves out the gray parts in b), c) and d), this violates some of the required properties for junctions. For details see text.

turn that also satisfies the *total reachability* assumption. But because neighboring exits and entries collapse, this subgraph violates the discreteness property. Even if one adds the two entries (En1, En3) and exits (Ex4, Ex2) like in Figure 7.7 d, one will not yet have identified a proper junction for the same reasons. Only if one adds the two channels at the center of the bidirectional road (Figure 7.7 a), this assumption is met.

We furthermore need *minimality assumptions* for junctions. This is because one can usually supplement a junction with further channels such that it satisfies the first two properties. In Figure 7.8 a, I have extended the median u-turn at the left end by completing the dual carriageway road. At the right end, I have added two channels. Junctions are required to be *minimal* in the following two senses: In a first sense, a junction is minimal because it *never contains a smaller junction.* I imply a criterion of individuation based on minimality, because junctions keep the first two properties if one adds an extra road to them (Figure 7.8). In a second sense, I require a *minimal vertex degree* for entries and exits: This is because entries and exits should afford either a navigational choice or a path from another entry, or to another exit. In the first case, the entry has more than one successor, and the exit has more than one predecessor. In the second case,

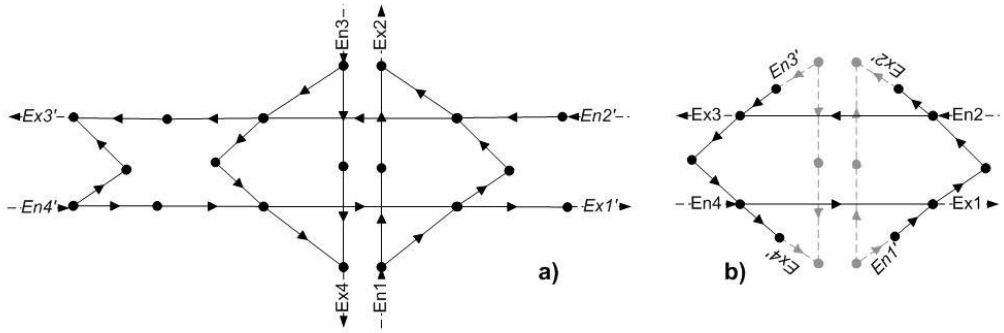


Figure 7.8: Illustration of the minimality properties by a *median u-turn* channel subgraph. a) It is always possible to add roads (left) and superfluous channels (right) and still meet requirements one and two for junctions. b) This subgraph even meets the first minimality requirement. For details see text.

the entry has an internal predecessor, and the exit has an internal successor. Why is this a legitimate property of junctions? Take for example the entry En2' with cardinality one in Figure 7.8 a. All paths from this entry must cross the one edge incident to En2'. And because no other path (from another entry) uses this edge, the subgraph can safely be shortened by this edge while retaining its total reachability property. The same applies for exits.

The second minimality criterion is necessary because it is generally not implied by the first. To see this, one has to slightly increase the complexity of the median u-turn junction (see Figure 7.8 b) by including additional vertices into the arcs that allow right turns from and to the vertical bidirectional street. Properties one and two are satisfied by the subgraph with full black arcs, because all entries and exits are gates and mutually reachable from each other. This subgraph is also minimal in the first sense, because it does not contain a smaller junction satisfying the properties one and two. But the minimal vertex degree of these new gates is one, so that this condition is able to identify the false positive example.

I use an abbreviation for the fact that in a pair of channels, one channel leads to the other:

Definition 83 $LConn(x, y) \leftrightarrow LeadsTo(x, y) \vee LeadsTo(y, x)$

This leads to the following recursive definition of a junction as a minimal class of channels:

Definition 84 $Junction^{n-way}(x) \leftrightarrow \exists y.$
$(\forall en.Entry(en, x, y) \rightarrow \neg Exit(en, x, y)) \wedge$
$(\forall ex.Exit(ex, x, y) \rightarrow \neg Entry(ex, x, y)) \wedge$
[discreteness of navigational action]

$(\exists!^n ex.Exit(ex,x,y) \wedge \exists!^n en.Entry(en,x,y)) \wedge$
[there exist n entries and exits]
$(\forall ex.Exit(ex,x,y) \rightarrow \exists!^{(n-1)} en.Entry(en,x,y) \wedge (ex)ReachableFrom(en)) \wedge$
[total reachability]
$(\neg\exists z.Junction^{n-way}(z) \wedge PP(z,x)) \wedge$ *[minimality I]*
$(\forall e.(Entry(e,x,y) \vee Exit(e,x,y)) \rightarrow$
$(\exists c,d.c \neq d \wedge c \in x \wedge d \in x \wedge LConn(e,c) \wedge LConn(e,d)))$
[minimality II]

Junctions (n-way) are induced minimal subgraphs of a channel digraph which have exactly n entries and exists which never coincide, and whose vertex degree is at least two, such that every exit is reachable from $n-1$ entries.

7.4. Application to data

Junction types. I will now demonstrate that this definition is satisfied by instances of the most common junction types. The first example I examine is a simple *4-way intersection* (Figure 7.9). The subgraph consists of four entry and four exit gates without internal channels, and these are directly connected by LeadsTo edges. As the subgraph is the minimal model for a 4-way junction, it is minimal in the first sense, and because each gate is a vertex with three successors (resp. predecessors), it is also minimal in the second sense.

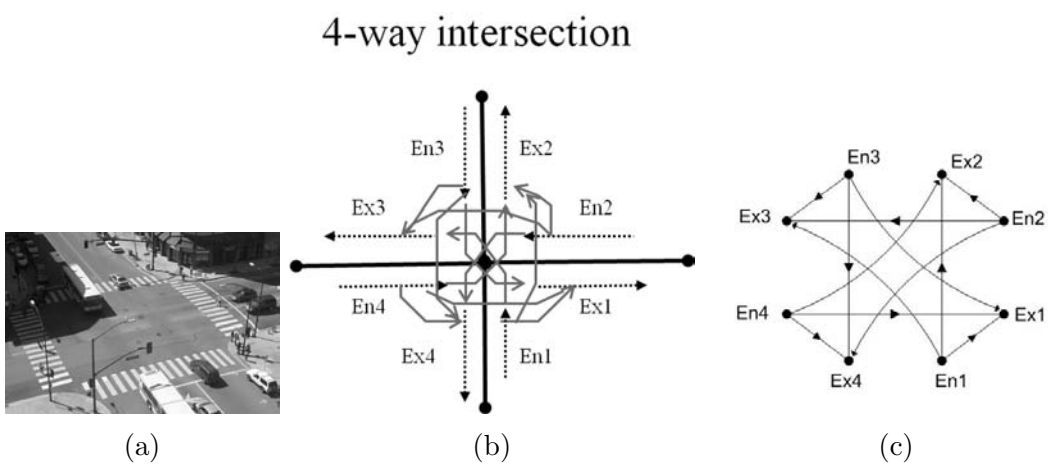


Figure 7.9: (a) A 4-way intersection (source: public domain), (b) its data model and (c) the corresponding channel digraph.

A junction type similar to the median u-turn is a *diamond interchange* (Figure 7.10). Here two major roads with different speed limits intersect. The long ramps indicate the faster road. The channels of the faster road are unnecessary for the individuation of a 4-way junction, therefore they are marked as feature

external (gray). As one can see all entries and exits are totally reachable gates with vertex degree two. The model is also minimal in the first sense: Although there is an *elementary circuit* due to the u-turns in the middle of the minor road, satisfying properties one, two, and four in Section 7.3.2, this circuit does not have gates. Any other totally 4-reachable subgraph does not comply with property four.

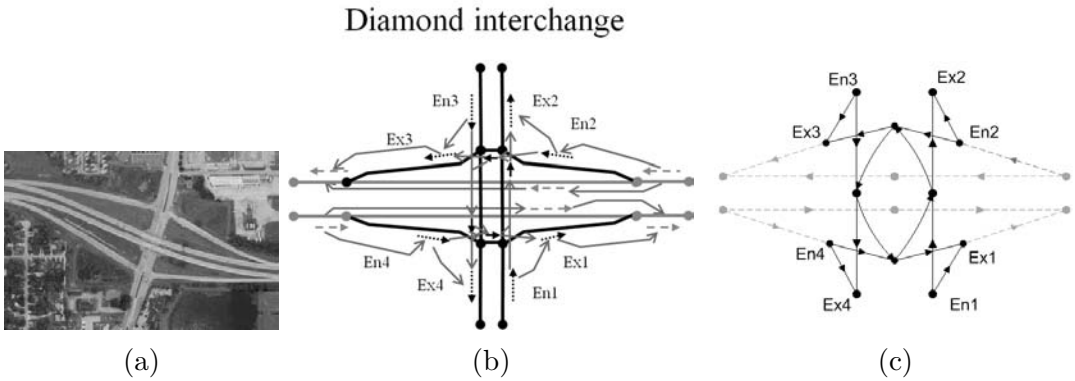


Figure 7.10: (a) A diamond interchange aerial photo (source: public domain (USGS)), (b) its data model and (c) channel digraph.

A 'prototype' junction is the *cloverleaf interchange* (Figure 7.11), mainly used for highways and high speed roads to prevent left turns by so called *loop roads*. The junction is totally reachable (it is possible to drive on an elementary circuit consisting of exactly those loops), and obviously has gates of degree two (properties one, two and four of Section 7.3.2). It is also minimal: although the inner circuit is totally reachable in four directions and all its vertices have a degree of two, they are not gates. All other subgraphs contradict one of the junction properties.

In addition to these models, our definition of a junction (Definition 84) is also satisfied by the following 4-way types: *quadrant roadway intersection*, *stacked interchange*, *continuous flow intersection*, and the following 3-way types: *trumpet interchange* and *semi-directional T-interchange*[8]. The considered junction types are deemed most common also by Wikipedia[9].

The question of whether our first-order logic definition is a *correct and complete classifier*, thus a solution to an inductive prediction learning problem covering all and only the positive examples [208], turns out to be problematic from a methodological point of view. First, because the concept of a junction and other higher order network features is not unambiguously established in existing network data models, it is unclear where to get an *unbiased sample* from,

[8]See [146] for an extensive list of signalized junction types with description.
[9]`http://en.wikipedia.org/wiki/Junction_(traffic)`

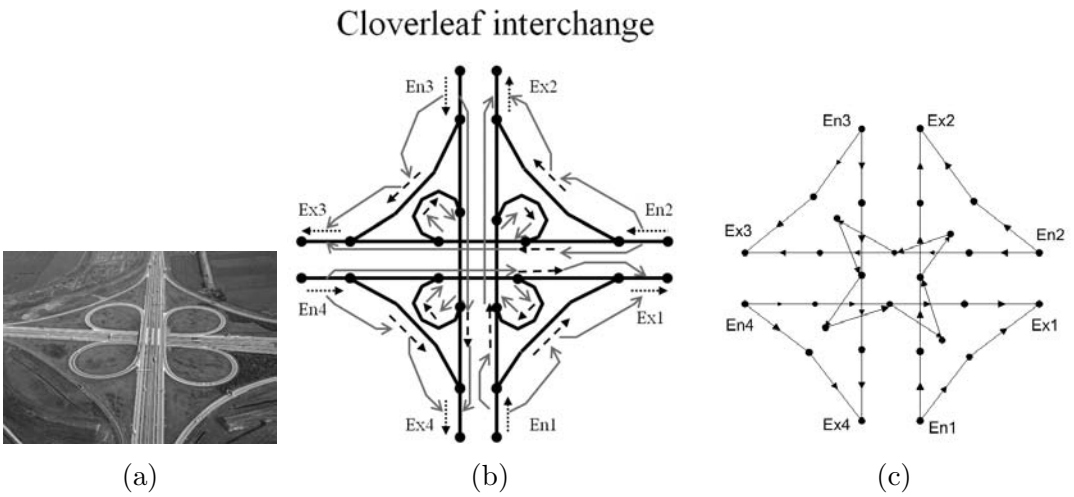


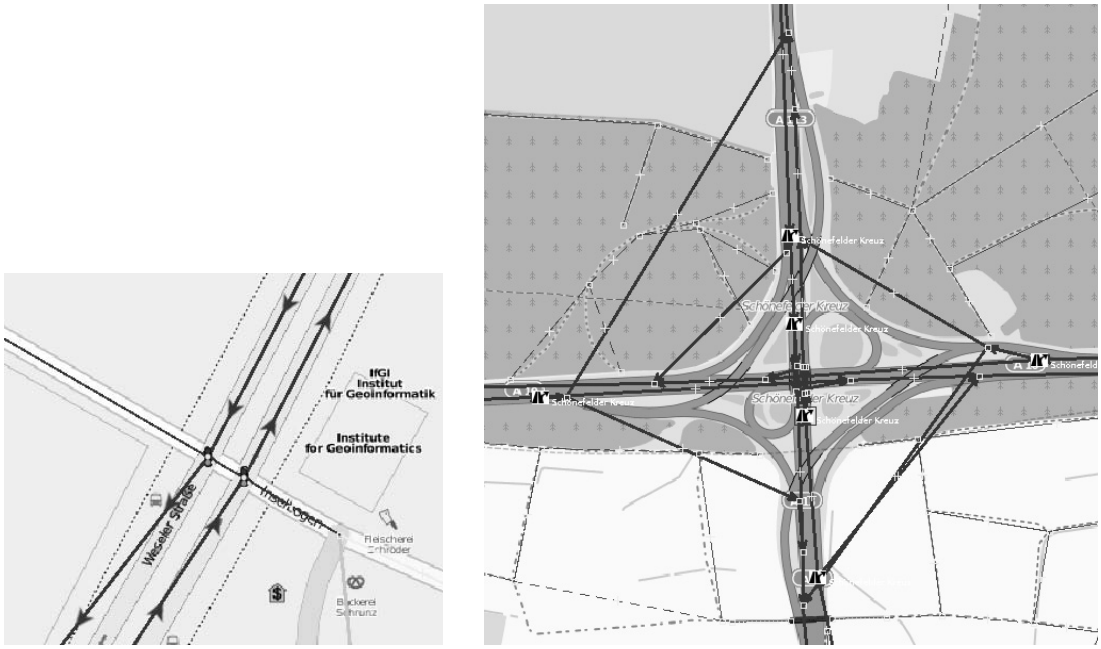
Figure 7.11: (a) A cloverleaf interchange (by kind permission of Rijkswaterstaat, Adviesdienst Geo-informatie en ICT ©, Beeldbank VenW, Netherlands), (b) its data model and (c) channel digraph.

and who applies the 'right' concept. Second, road network categories—like all human categories—have a *graded structure* [90], so there are many more or less prototypical versions of a definition. Our junction definition is also applicable to *roundabouts* that have only gates as entries, though for a general definition of roundabouts (which can include non-gate entries), the discreteness property of Definition 84 would have to be dropped. Thus, roundabouts are less prototypical types of junctions needing special treatment.

Search for junctions in OSM data. The proposed affordance-based junction definition has been turned into a depth-first search graph algorithm [156] and applied to OSM datasets [129]. Compare [156] for details on how to translate the recursive Definition 84 into an algorithm. In the latter work, we proposed how to interpret the channel network theory into OSM tags, nodes and ways, and implemented a junction search tool[10]. I will present here some result of applying the junction search tool to OSM in oder to illustrate the practical use of a reference theory, but refer to the cited works for further details.

The tool allows to search for n-way junctions where n is a parameter to be chosen by the user. It first requires to generate a channel digraph from an ordinary OSM road network, where the *LeadsTo* relation among channels is inferred from available OSM tags. Unfortunately, this can only be an approximate heuristic since OSM data is incomplete with respect to turn restrictions [129]. This is also the reason why certain junctions cannot be properly individuated. Then the user can select a subset of channels (an induced subgraph) to be scanned for junction

[10]See `http://wiki.openstreetmap.org/wiki/JOSM/Plugins/JunctionChecking`.



(a) (b)

Figure 7.12: (a) Junction "Weselerstr./Inselbogen" in Münster. (b) Intersection "Schönfelder Kreuz" near Berlin. Dark arrows depict channels (source: OSM).

instances of the given type.

The first example is a rather simple 4-way junction with a dual carriageway, which was identified by the tool as a set of channels highlighted in dark color (Figure 7.12). Turn restrictions in this case can be correctly inferred from the information that the dual carriageway is represented by two one-way streets. A more complicated 4-way junction is the highway intersection "Schönfelder Kreuz" near Berlin (Figure 7.12), consisting of 47 channels highlighted in dark. The identification is correct since turn restrictions can be inferred from the configuration of one-way streets. Channels are depicted as simplified directed arcs in the map where each channel is interpreted into one or several OSM ways[11]. Even highly complicated junctions can be identified, as the 5-way highway junction "Wuppertal Nord" consisting of 71 channels (Figure 7.13). For further examples and a discussion of the limitations, see [156].

7.5. Constructible road network categories

In this chapter, I have illustrated how essential semantic categories underlying road network data can be reconstructed based on the approach proposed in this

[11]This depends on the existence of topologically unnecessary helper nodes.

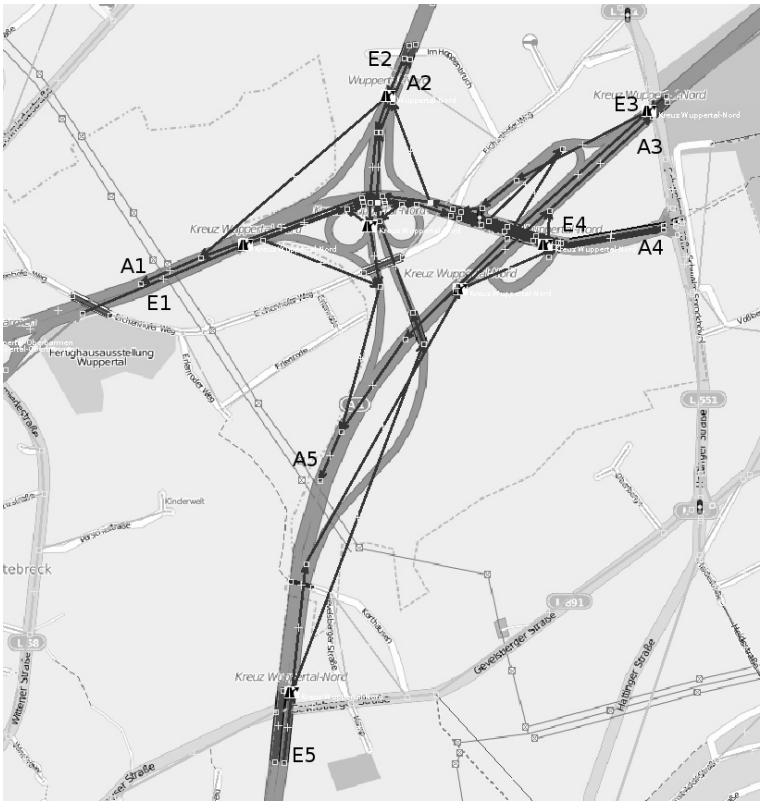


Figure 7.13: Highway junction "Wuppertal Nord" is an example for a 5-way junction that is correctly identified in OSM (source: OSM).

thesis. I started from a particular kind of social affordance, called turn-compliant movement affordance, which is a subtype of a supported locomotion affordance. Based on the idea that such locomotions can be simulated in "bundles" on a road as long as there does not exist a turn off possibility (which means the observer is not able to simulate a locomotion that breaks the bundle), I defined the notion of a channel. I also defined an affordance relation between channels in terms of connecting locomotion affordances.

These two defined notions, which were taken as primitive in [154], have corresponding equivalents in a routable road network dataset: a channel usually corresponds to a direction of an embedded road segment (e.g., of a non-dissected part of a 'way' in OSM), and the *LeadsTo* relation corresponds to the set of consecutive pairs of channels that are not affected by turn prohibitions. The two notions can also be used to define channel networks, which are fully reachable wholes of channels. These stand for complete road networks.

Furthermore, I defined n-way junctions, like in [154], as induced minimal subgraphs of a channel network with n discrete entries and exits of vertex degree two, and in which every exit is reachable from $n-1$ entries. I illustrated that

Name	Formal symbol	Definition
channel	*Channel*	76
connection of channels	*LeadsTo*	77
reachability	*ReachableFrom*	78
channel network	*ChannelNetwork*	79
entry of induced c.-subgraph	*Entry*	80
exit of induced c.-subgraph	*Exit*	81
junction	$Junction^{n-way}$	84

Table 7.1: Constructible road network categories.

this definition is satisfied by a collection of junction types, and shortly discussed its implementation and test in a junction search tool for OSM. The constructed notions are summarized in Table 7.1.

What is the practical benefit of the proposed constructions? First, they allow us to become aware of the *roots of reference of a concrete data set*, i.e., how its asserted facts may be observed and how its categories should be individuated. This is not at all a clear matter. What makes a road a road? One might think of roads as rigid physical objects. But consider the infrastructure in developing countries: they are not generally identifiable by paved surfaces or signs. Paved surfaces do not make a road, as little as a backrest makes a seat. But even in an environment with clearly demarcated roads, it remains unclear where a junction such as "Wuppertal Nord" begins and where it ends.

Second, it becomes now possible to *compare* different data sets according to *which kinds of observations they represent*, and which they are lacking. As Possin argued [129], OSM data – in contrast to commercial road datasets – is not routable in large parts, because the necessary information about turn restrictions is incomplete or lacking. Based on previous work, turn-compliant locomotion affordances were identified as a central kind of observation that underlies the concept of road categories.

Third, it becomes possible to *compare data sets of different granularity*. Geographic information needs to be represented at different levels of granularity. But how to map these different levels, and how to generalize maps in general, is an unsolved problem [157]. Complex road features such as junctions may be represented as a single node, as well as a complex subgraph, or even as a connected surface in a raster map. The method proposed in this thesis can account for all these levels, because it constructs categories from the bottom up.

Fourth, since a reference theory allows for logical inference, it becomes possible to *enrich data sets*. For example, if one finds *graveyards* and *factories* in a road network graph, one can infer that there is an error or a missing affordance link in the data set. The proposed junction search tool can be used to add category labels and junction relations to OSM.

Chapter 8

Conclusion

Was sich in der Sprache spiegelt, kann sie nicht darstellen.
Was sich *in der Sprache ausdrückt, können* wir *nicht durch sie ausdrücken.*

— L. Wittgenstein [206]

Following Wittgenstein's insight above, I posit that a data semantician has to accept that the things which express themselves in language are only partially expressible in it. Therefore, any solution to the problems of reference and grounding in data semantics has to rely on some machinery that goes beyond language. The challenge is to say concretely what this machinery consists in and how we can refer to it in a way that is reproducible, so that language interpretations of provider and user of a data set can be coordinated.

In this thesis, I have argued that current semantic strategies are challenged by this requirement, because the problem of reference is not solvable by any of them alone (Chapter 2).

I have then discussed a human attentional apparatus (Chapter 3) that may solve the problem based on perceptual operations: A focus of attention is a memorized discrete moment in the mind of an observer that is produced by *focusing attention* on some mental domain, e.g., on the space around the body (ISR). This space is a pre-conceptual mental product of the human brain in which Gestalts are precomputed on sensory-motor inputs. Observers experience their environment by applying these Gestalts and memorizing their presence at attentional foci. The relevance of this mechanism lies in the fact that human attention can be joined among observers in a common scene, and thus allows people to *refer* to the Gestalts and to *predicate* their presence from a certain perspective. These operations are called *referencing* and *predication*, and are parts of shared speech acts. As such, they constitute lowest level *information* items. Predications in the ISR domain are also called perceptual operations, since they draw on perceptual mechanisms. Furthermore, the meaning of speech acts also involves constructive operations. These are used to create abstract entities we can refer to with language.

Chapter 4 treated the problem how a formal technical language can be established based on this apparatus, such that its primitives are tied to it and mental constructions can be guided by it. I have argued with Quine, Piaget and von Glasersfeld that abstractions can be conceived as logic reifications that are computed on represented actions. The memorized relations on foci of attention are results of perceptual operations. They represent particular action performances, and as such can be the input to abstractions. Furthermore, action types might be represented by recursive rules (following Lorenzen) or singular entities. I have then proposed that Quine's idea of observation sentences can be used in order to conventionally establish a technical language, in particular, to express predications by applying *observation predicates*, and to construct entities with the help of *existential quantification*. In consequence, such an observation language has an ordinary Tarskian semantics in terms of sets of abstract entities together with foci of attention in a momentary window of attention, as well as an operational meaning in terms of underlying perceptual operations. The expressivity of this grounded language is ordinary FOL, and since it has finite models, it can be used to recursively define notions, such as transitive closures, and to individuate extensional reifications, such as unified wholes. Furthermore, I suggested a non-extensional form of abstraction called analogical reification, which accounts for the existence of non-observed entities in reference systems, e.g. calendars, and for the existence of infinity abstractions in mathematics, and which may be based on recursive rule representations of actions.

In Chapter 5, I have made proposals for observation predicates, such as $\leq_T$ for the temporal ordering of the domain F of attentional moments, SC for the connectedness of foci by a visual surface in the environment, PF for predicating point-like features with respect to a surface host, such as one end of a bar, OnL and $=_L$ for predicating linearity and the comparison of lengths among foci, DoX for identifying actions, and $AffordsX$ for predicating the presence of action affordances. A reference theory is an axiomatized grounded FOL theory that has such predicates as primitives and the intended finite universe of interpretation D, which consists of attentional moments F and all abstractions.

In Chapter 6, I have proposed a formalization of a reference theory that deals with basic phenomena of our immediate environment relevant for grounding geodata. I have proposed a finitist experiential geometry of locations, constructed relative to a spatial reference frame made of point-like features. I constructed the notion of angles and length from the observation predicates OnL and $=_L$. Based on these geometric primitives, I proposed a finite notion of strong connectedness. I then presented formalizations of Gibson's meaningful environment, such as media of illumination and locomotion, substances and surfaces, air and ground, in terms of locomotion, breathing and optical affordance. These notions can be used to define meaningful properties, such as the waterdepth of a waterbody, and meaningful media such as roads and places.

Chapter 7 picked up the thread and illustrated how a common geographic

data model, namely the one of a road network database, can be grounded in this reference theory. By means of road network junctions, I demonstrated how an affordance based definition can help to categorize and identify complex geoobjects. The definition was evaluated and implemented in a junction search tool for Open Street Map.

Reference theories like the one proposed here can help us understand a data set in terms of its roots of reference, i.e. in terms of *what needs to be observed and constructed in order to obtain it*. This includes, in particular, what needs to be predicated about an environment. The underlying perceptual operations are thereby undeceivable in the sense that they need to be performed, not described, in order to solve the grounding problem of a formal language. This is why the proposed method can practically rely on their results without describing how they work. This of course does not mean that it may not be insightful to describe them, which could only be done in a preliminary way in this thesis, and may be more a task for cognitive science than information semantics.

From a practical viewpoint, reference theories can be used to *establish a shared domain vocabulary* that describes a data set. They therefore can play the role of a foundational ontology. But they may also be used to *ground abstract terms in these ontologies* themselves. They seem to be promising candidates for a general *ontological backbone* that allows to compare different vocabularies with respect to their ontological commitments. Since the proposed meta-theory accounts for constructive freedom, it may even be used to explain how it comes that different foundational ontologies express different and formally incompatible views on the same phenomenon. For example in order to compare nominalist or reductionist with realist or multiplicative standpoints about qualities [2]. I suppose that many of the essential criteria for ontological soundness identified by Guarino and Welty [58] may be given a grounded interpretation. This, of course, is not to say that reference theories could in any way substitute current information ontologies.

Furthermore, since reference theories are logical theories, they allow for inference and thus to *enrich data sets* that were annotated with it. They do not differ in this respect from information ontologies.

The reference theory proposed is a first order theory and thus computationally intractable. It may therefore not directly be used to annotate a data set. Its role is more to *shape and compare light weight vocabularies* in an ontologically sound way. For example, the design of OSM tags for annotating points of interest can be based on the theory of affordances proposed in this thesis, as proposed in [160]. Similarly, sensor data may be described by their underlying observation procedures using a reference theory [160], and semantic web vocabularies may be compared based on how the annotated data sets were obtained by a human observer. For this purpose it is necessary to translate the reference theory into a tractable language like OWL[1].

[1] 'Web Ontology Language', see `http://www.w3.org/TR/owl2-overview/`.

As illustrated by the junction category in Section 7.5, grounded reference theories seem particularly suitable to *compare data sets of different granularity*. Geodata, and especially road network data, are paramount examples of how different resolution of the same situation is required but still poses problems of generalization.

Based on these research opportunities, I consider the following as preferable future work:

- *Grounding as a backbone for central ontological distinctions.* Using ontological distinctions like identity and unity criteria, Guarino and others have developed a framework for constructing sound ontologies [59]. I have already shown that reference theories are capable of delivering identity and unity criteria for ontologies. It may be worth investigating how it can explain and amend such ontological frameworks through reconstruction from perceptual operations. For example, one might distinguish between sortals (these are categories 'carrying' identity criteria, e.g. 'being human' or 'being a color', where one category instance can be distinguished from the other using the underlying concept), and non-sortals (categories like the universal quality value 'redness', whose instantiations cannot be distinguished by the underlying concept) on the basis of their constituting perceptual operations. Furthermore, many useful ontological distinctions (compare the ideas presented in [109]) await grounding. For example, the distinction between perdurants and endurants, the construction of universals [118], the construction of processes and events [46], as well as the construction of amounts of matter [60, 160]. An important opportunity for grounding are also social constructions, as described in [166].
- *Investigate the cognitive properties of operations for referencing and predication.* The proposed predications have a speculative character even though they are motivated and justified by cognitive research. Pinpoint cognitive investigations such as [79] could help to sharpen our understanding of them, in particular, of what can be done and cannot be done with them.
- *Extend the reference theory to cover semantic sensor concepts* [170]. The goal is to demonstrate that the approach can be used as a basic method for annotating sensors and measurements. Since technical sensors need to be calibrated to be interpretable, and calibration is based on human observation procedures, a grounding approach based on such procedures suggests itself.
- *Extend the reference theory to cover place concepts.* Practically relevant location concepts and especially concepts of place go beyond geometrical primitives [29]. I expect that places can be conceived as kinds of media grounded in affordance primitives [152]. Take for example Johnson's image schemas [74] like 'path' and 'container', where a container, e.g. a house, affords shelter to persons. Or Kevin Lynch's [101] imaginable elements of

a city, like 'path', 'edge' and 'district'. Lynch's 'edge', for instance, which can be a railroad cut, a wall, a shore etc., can be conceived in terms of the absence of a walking affordance in a city, while a 'path' in terms of its presence. As a first step, it would be valuable to develop a model of different types of actions relevant for *a selected domain of POIs*, e.g. POIs related to eating and drinking in OSM.

- *Ground the GDF standard for road networks.* GDF[2] is the current data standard for road networks. It contains specifications for features ('road element', 'form of way'), attributes and relationships. But it is still unclear in many respects and allows for incompatible interpretations. Together with proprietary content extensions, this causes lacking interoperability among commercial datasets. The road network theory proposed in this thesis suggests itself in order to propose more precise specifications or to highlight and analyze incompatibilities among data sets, since it is based on primitives that are independent of any specific road conceptualization.
- *Generate a "light-weight" reference theory in OWL or some rule based web language.* In order to be useful for annotation and query in the semantic web, the grounded notions defined in a reference theory need to be translated into some less expressive language.
- *Investigate further constructive formalisms.* In this thesis, I have used FOL theories as language standard for reference theories. These express operations such as predications and referencing only in terms of their memorized results. As I argued in Section 4.1.2, and as I have suggested in [155], an alternative is to represent the operation types themselves in terms of recursive rules in a proto-logic [98], and logic operators in terms of dialogue games [100]. This has two advantages. First, through such languages, it becomes possible to generate a domain of discourse D, instead of presuming it in a model. And furthermore, as I argued in Sections 4.1.2, there are certain kinds of abstractions which seem to require a recursive representation of an operation (see next bullet point).
- *Investigate non-extensional logical reifications and fictions.* There is a large variety of abstractions which cannot be captured by extensional class abstractions, including non-observed entities in reference systems as well as abstractions based on metaphors. In Section 4.2.2, I made informal suggestions to capture these in terms of what I called *analogical reifications.* These are indispensable for semantic reference systems, but still need to be analyzed and spelled out. An encouraging possibility is to base them on recursive rules representations.

  Furthermore, there is a variety of abstractions that were proposed by Vaihinger in [197] which await discovery. One useful technical fiction is, e.g.,

[2] 'Geographic Data Files'. In the newest version GDF 5.0, see `http://en.wikipedia.org/wiki/Geographic_Data_Files`.

a statistical population, which is a frequent term needed to understand the semantics of statistical parameters.

- *Ground the notions 'resolution', 'granularity' and 'uncertainty'.* According to Frank [43, 42], observations constitute the basic information layer for spatio-temporal databases, and therefore resolution is a result of a certain observation process. If one follows Frank, the grounding method proposed here suggests itself in order to provide a deep analysis of the notions of granularity, resolution and uncertainty. One could expect to arrive at a more adequate account of data quality than provided by current theories. Common notions of data quality need grounding themselves, as the idea of an error measured with respect to "reality" is a pure fiction. At the end of the day, quality assessments must be constructs based on comparisons of different kinds of observations.

The approach followed in this thesis might be promising for many branches of contemporary information science. However, for most of these areas, its value still needs to be demonstrated. Also, its complexity is a challenge for understanding. One might argue that a grounded reference theory deviates too far from understandable common sense in order to be actually used.

In response, I admit that reference theories are complex, but argue not to underestimate the complexity which underlies ordinary language comprehension, as well as our abilities to cope with it. In reconstructing Lorenzen's linguistic vessel on the sea[3], we cannot expect an auxiliary boat and must begin as swimmers. But in doing so, we can rely on our embodied competence of language comprehension. Reconstruction is a challenge that is mastered by us every time we learn a new word. And reference theories just do what we as competent speakers have already done multiple times: they refer to what we want to say in case our ordinary language fails.

[3]Compare [100] and Section 3.4.

Appendix

8.1. The difficulties of learning references in the Chinese Room

The basic insight is that the consistency of a theory with empirical evidence never rules out the possibility that there is more than one way of translating its names and predicates into another theory which is satisfied by the same empirical facts.

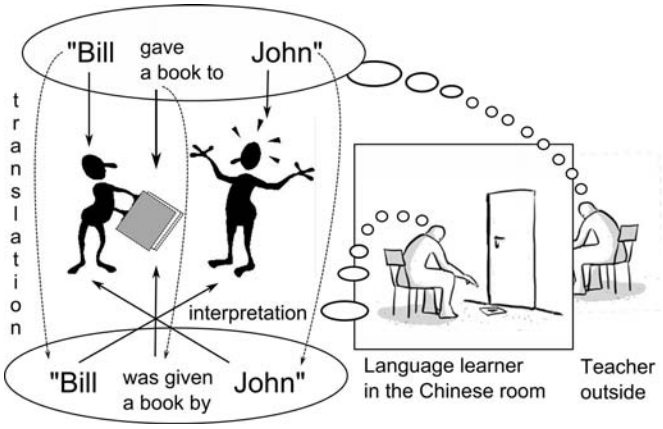


Figure 8.1: Illustration of the difficulties of learning language references in the Chinese Room. The Chinese Room illustration was adopted from `http://en.wikipedia.org/wiki/Chinese\_room`.

What this means is that the extension of a sentence like "Bill gave the book to John" can be isomorphically mapped into the same (or a one-to-one corresponding) domain, such that the entity denoted by "Bill" is mapped to the entity denoted by "John" and vice versa, and that the relation denoting "gave the book to" is mapped to the relation "was given a book by" (Figure 8.1). While the resulting interpretation still satisfies the sentence, it is now totally different.

In order to illustrate the difficulty of learning referencing and predication under these circumstances, let us assume a modification of Searle's Chinese Room setting (compare Figure 8.1): A person receives written sentences like the one above in an unknown language, e.g. Chinese, from a language teacher outside of

a closed room. Suppose that teacher and learner have had acquaintance with the same perceivable scenes (they both know John and Bill, and they know that Bill gave the book to John), but that they neither share a common language nor were both present in any commonly perceived scene, in which one could have indicated the reference by pointing. Suppose that the second interpretation is the result of the learner's interpretation of the teacher's sentences uttered with interpretation one in mind. Then a definite description like "the person that gave the book to John", uttered by the teacher in his or her own language, could be interpreted by the learner as John (in case the mapping goes wrong as indicated above) or Bill (in case the mapping happens to be right). In effect, inside the Chinese Room, it would be impossible to learn the correct predication as well as referencing of a language, since they cannot be distinguished on the basis of their syntax.

8.2. Proof of the triangle inequality

In addition to theorems in Sections 6.2 and 6.1, we need one further theorem in order to prove the triangle inequality. This theorem is constructive and states that for any triangle abc with a point a^* between a and b (and a certain configuration of auxiliary points), the angle cab must be smaller than angle ca^*b[4] (compare Figure 8.2). We first have to define a certain *auxiliary point configuration* on this triangle called $ARPC$:

Definition 85 *Angle Reflection Auxiliary Point Configuration:*
$(a, b, c, a^*)ARPC(c_1, c_{11}, c_{12}, c_2, m, x, d, d^*) \leftrightarrow a \cong_b a^* \wedge$
$Rf(c, a^*, c_1) \wedge OnL(a, a^*, c_{11}) \wedge Rf(c_{11}, a^*, c_{12}) \wedge ca^* =_L a^*c_{11} \wedge$
$Rf(a, m, a^*) \wedge Rf(c, m, c_2) \wedge Rf(d, m, d^*) \wedge ma^* =_L a^*d \wedge$
$OnL(m, x, c_1) \wedge OnL(a^*, x, c_2)$

In this construction, it is always the case that $cam \leq_A ca^*c_{11}$ for some point a^* lying on the ray starting from b through a (Figure 8.2). This result is obtained by reflecting those angles onto each other by way of parallelograms. Additionally, if it is the case that a^* is a point between ab inside the triangle, the construction can be used to compare angles ca^*b and cab, because they correspond to ca^*c_{11} and cam (compare Figure 8.2). We first reflect point c two times, once in point a^* and once in the midpoint m between aa^*. Then we construct a rectangle c, c_1, c_{11}, c_{12} which allows us to reflect angle ca^*b onto angle $c_1a^*c_{12}$. The rectangle a, d^*, a^*, d allows to reflect angle cab onto angle ma^*c_2. In order to compare the two reflected angles, we only need some additional auxiliary point x. For this configuration and the assumption $OnL(a, a^*, b)$, we can easily prove that cab must be smaller than angle ca^*b:

[4]This is a specialisation of the more general theorem that outer angles are always bigger than the two inner angles on the opposite side of a triangle, compare [162], sentence 11.41. The proof given there is more involved regarding the constructive effort needed.

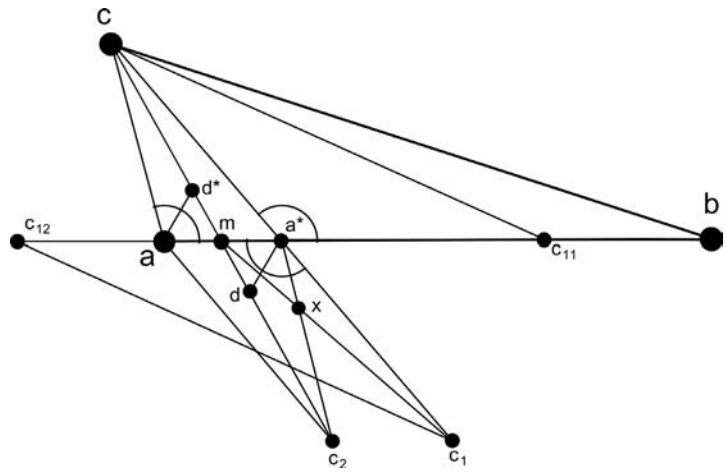


Figure 8.2: The auxiliary point configuration $ARPC$ needed to prove that $cab \leq_A ca^*b$ if $OnL(a, a^*, b)$.

Theorem 24 *Theorem about angles at inner triangle points:*
$(a, b, c, a^*)ARPC(c_1, c_{11}, c_{12}, c_2, m, x, d, d^*) \wedge OnL(a, a^*, b) \rightarrow cab \leq_A ca^*b$

Proof Using Theorem 11, we can infer that $c_1c_{12} =_L c_{11}c$ in the first rectangle, and thus by Definition 30, $c_1a^*c_{12} =_A ca^*c_{11}$. Since $OnL(a, a^*, b)$ by condition and $OnL(a, a^*, c_{11})$ by Definition 85, we can infer $ca^*b =_A ca^*c_{11}$ by applying Definition 30 and using connectivity Axiom 21 of OnL. We then get $c_1a^*c_{12} =_A ca^*b$ by transitivity (1). Similarly for the other rectangle, by Theorem 12 we get $ac =_L a^*c_2$, and since $cm =_L mc_2 \wedge am =_L ma^*$ by construction, $ma^*c_2 =_A cam$ by Definition 30. Since $OnL(a, a^*, b)$, we can infer also $cam =_A cab$ by applying the same definition and using connectivity of OnL (Axiom 21). So we have $ma^*c_2 =_A cab$ by transitivity of angle congruence (2). Since by condition, $OnL(m, x, c_1)$ and $OnL(a^*, x, c_2)$, we can apply Definition 31 and get $ma^*c_2 \leq_A c_1a^*m$, and by Definition 30 also $c_1a^*c_{12} =_A c_1a^*m$ (3). By transitivity we can combine results (1), (2), (3) to get the required result $cab \leq_A ca^*b$. *q.e.d.*

We are now ready to prove the triangle inequality in existentially conditioned form, i.e., as a constructive theorem. The prove is by contradiction. We simply disprove that it is possible to have point a^* lying inside the segment ab if the length of ba^* corresponds to the non-straight path b, c, a in the triangle (compare Figure 6.4). The required auxiliary point configuration is included as existential condition in terms of the following abbreviation:

Definition 86 *Auxiliary points needed to prove the triangle inequality for abc:*
$\forall a, b, c.APTI(a, b, c) \leftrightarrow$
$[\exists c^*.OnL(c^*, c, b) \wedge ac =_L cc^* \wedge [\exists a^*, c_1, c_{11}, c_{12}, c_2, m, x, d, d^*.$
$a^*b =_L c^*b \wedge (a, b, c^*, a^*)ARPC(c_1, c_{11}, c_{12}, c_2, m, x, d, d^*)]]$

Theorem 25 *Triangle Inequality (Constructive):*
$\forall a, b, c.APTI(a, b, c) \rightarrow ab \leq_L bc + ca$

PROOF From the first row of the condition and by Theorem 14, we know that $bc^*a \leq_A c^*ab$ (1). Since $a^*b =_L c^*b$ by the second row of the condition, a^*c^*b is an isosceles triangle and we get $bc^*a^* =_A c^*a^*b$ by Theorem 13 (2). Now we proof by contradiction that $OnL(a^*, a, b)$: Suppose the contrary, $\neg OnL(a^*, a, b)$. But this must mean that $OnL(a, a^*, b) \wedge \neg a =_{Ref}^{S,A_1,A_2,A_3} a^*$, since by the *ARPC* condition and Definition 85, we know that $a \cong_b a^*$. It immediately follows from Definition 31, that $bc^*a^* \leq_A bc^*a$. Since additionally $\neg a =_{Ref}^{S,A_1,A_2,A_3} a^*$, it follows that $\neg bc^*a^* <_A bc^*a$. By transitivity of angle congruence, we can combine this result with result (2) to get $c^*a^*b <_A bc^*a$. And by transitive application of result (1) to this result, we get $c^*a^*b <_A c^*ab$. But from Theorem 24 and the *ARPC* condition, we know that $c^*ab \leq_A c^*a^*b$, which would result in a contradiction. So it must be the case that $OnL(a^*, a, b)$ (3). It immediately follows $ab \leq_L a^*b$ by Definition 25, and since $a^*b =_L c^*b$, it follows that $ab \leq_L bc + ca$ by Definition 27. *q.e.d.*

8.3. Proof of consistency and incompleteness of the proposed reference theory

By a model of a reference theory I understand a structure $\mathfrak{M} = \langle A, R_1, .., R_n \rangle$, where A is a nonempty set called domain and $R_1, .., R_n$ are relations on this set, and there is an interpretation of the theory's observation predicates in terms of these relations such that all axioms of the theory are satisfied.

In the following I will describe two such models, based on finite subsets of the *4-dimensional vector space over the real numbers*, $\mathfrak{C}_4(\mathbb{R}) = \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$. Each point $\vec{x}$ in this space is given in terms of a vector of dimension indexed symbols $\vec{x} = (x_1, x_2, x_3, x_4)$, where the constant $\vec{0} = (0, 0, 0, 0)$. The usual vector operations are available on these points. The first three dimensions of each vector manifest space and the fourth one time.

The first model $\mathfrak{M}_I$ has the following domain:

$$A_I = \Big\{\vec{0}, \vec{a}^* = (-1, 0, 0, 2),$$
$$\vec{a}_1 = (1, 0, 0, 1), \vec{a}_2 = (0, 1, 0, 3), \vec{a}_3 = (0, 0, 1, 4),$$
$$\vec{a'}_1 = (2, 0, 0, 5), \vec{a'}_2 = (0, 2, 0, 6), \vec{a'}_3 = (0, 0, 2, 7),$$
$$m_1 = \left\{\vec{a}_1, \vec{a'}_1\right\}, m_2 = \left\{\vec{a}_2, \vec{a'}_2\right\}, m_2 = \left\{\vec{a}_3, \vec{a'}_3\right\}\Big\}$$

Observation predicates[5] are interpreted into the following defined relations on $\mathfrak{C}_4(\mathbb{R})$ with corresponding subscript:

Definitions

$R_{\leq_T}(\vec{x}, \vec{y}) \leftrightarrow x_4 \leq y_4$

$R_F(\vec{x}) \leftrightarrow \vec{x} \in \left\{\vec{0}, \vec{a}^*, \vec{a}_1, \vec{a'}_1, \vec{a}_2, \vec{a'}_2, \vec{a}_3, \vec{a'}_3\right\}$

$R_S(\vec{x}) \leftrightarrow \exists i \in \mathbb{R}.\vec{x} = (0, 0, 0, i)$

$R_{A_1}(\vec{x}) \leftrightarrow \exists i \in \mathbb{R}.\vec{x} = (1, 0, 0, i)$

[5]Note that an observation predicate is a constant, therefore the meta-variables DoX, XF, *AffordsMoving*, *AffordsX* used in the Axiom Schemes become observation predicates only if they were instantiated by some concrete type.

$R_{A_2}(\vec{x}) \leftrightarrow \exists i \in \mathbb{R}.\vec{x} = (0,1,0,i)$
$R_{A_3}(\vec{x}) \leftrightarrow \exists i \in \mathbb{R}.\vec{x} = (0,0,1,i)$
$R_{RefFrame}(\vec{x},\vec{x}_1,\vec{x}_2,\vec{x}_3) \leftrightarrow R_S(\vec{x}) \wedge R_{A_1}(\vec{x}_1) \wedge R_{A_2}(\vec{x}_2) \wedge R_{A_3}(\vec{x}_3)$
$\vec{x}\vec{y}R_{=_L}\vec{v}\vec{w} \leftrightarrow \sum_{i=1}^{3}(x_i - y_i)^2 = \sum_{i=1}^{3}(v_i - w_i)^2$
$\vec{x}R_{=_{Ref}}\vec{y} \leftrightarrow x_1 = y_1 \wedge x_2 = y_2 \wedge x_3 = y_3$
$R_{SC}(\vec{x},\vec{y}) \leftrightarrow \vec{x}R_{=_{Ref}}\vec{y} \wedge \left(\vec{x}R_{=_{Ref}}\vec{a}_1 \vee \vec{x}R_{=_{Ref}}\vec{a}_2 \vee \vec{x}R_{=_{Ref}}\vec{a}_3\right)$
$R_{PF}(\vec{z},\vec{x},\vec{y}) \leftrightarrow \vec{z} = \vec{x} \wedge R_{SC}(\vec{x},\vec{y})$
$R_{VertAln}(\vec{x},\vec{y}) \leftrightarrow x_1 = y_1 \wedge x_2 = y_2$
$R_{OnL}(\vec{x},\vec{y},\vec{z}) \leftrightarrow \exists r \in \mathbb{R}.\vec{x} + r \cdot (\vec{z} - \vec{x}) = \vec{y} \wedge 0 \leq r \leq 1$
$\vec{x}\vec{y}R_{\leq_L}\vec{v}\vec{w} \leftrightarrow \exists \vec{u}.R_{OnL}(\vec{v},\vec{u},\vec{w}) \wedge \vec{x}\vec{y}R_{=_L}\vec{v}\vec{u}$
$R_{AffordsSeeing}(\vec{z},\vec{x},\vec{y}) \leftrightarrow \vec{0}\vec{a}_1R_{\leq_L}\vec{0}\vec{z} \wedge R_{OnL}(\vec{0},\vec{z},\vec{x}) \wedge R_{OnL}(\vec{0},\vec{z},\vec{y})$

This model basically consists of the 5 vectors that build a spatial reference system with S in the origin of the 3-dimensional subspace that manifests space. Note that each point has a different temporal value x_4. There are also 3 points $\vec{a'}_1, \vec{a'}_2, \vec{a'}_3$ which lie one step beyond but on a line with the three extremities of this spatial reference system. Two vectors are on the same location if they have equal scalar values with respect to these 3 spatial dimensions.

Equidistance is interpreted as Euclidean distance. Verticality is expressed as the direction of the 3rd dimension. Surface connected are all vectors that have the same location as either one of $\vec{a}_1, \vec{a}_2, \vec{a}_3$. Those vectors are at the same time also on a point-like feature, and so surfaces are point-like in our model. Visible are all vectors that lie on the boundary or outside the sphere centered on the origin and having radius $\vec{0}\vec{a}_1$. OnL is expressed in the obvious way in terms of a linear vector equation.

Furthermore, the domain contains 3 classes m_1, m_2, m_3 defined by enumeration. It can be easily seen that these classes are illuminated media, and therefore non-overlapping wholes: Consider the three optic spheres centered on $\vec{a'}_i$ and bounded by $\vec{a}_i$. Every sphere connects $\vec{a'}_i$ and $\vec{a}_i$ as an optic sphere chain, and so they are part of the same illuminated medium. Since there is no sphere chain connecting these spheres, they are unconnected and non-overlapping.

We interpret all remaining (action and affordance-based) observation predicates into the empty relation, i.e., into the empty set $\{\}$. Furthermore, we interpret the element of relation $\in$ into the relation between elements and subsets of $\mathfrak{C}_4(\mathbb{R})$. We now can prove consistency.

Theorem 26 *Consistency:*
$\mathfrak{M}_I$ is a model of the reference theory described by all Axioms and instantiated Axiom Schemas in Appendix 8.4.

Proof Axioms 1 and 2 are obvious truths of regular sets. In particular, we have in our model only a singular layer of sets, i.e., $n = 1$. Note that for each of the three classes m_1, m_2, m_3, we can easily define a predicate $\Psi_i(\vec{x}) \leftrightarrow R_{OnL}(\vec{0},\vec{a}_i,\vec{x})$ with $i = 1,2,3$ that accounts for its

instances. Therefore Axiom Schema 1 is satisfied. Since there are no further classes, Axiom Schema 2 is satisfied, too. Since F consists of vectors from $\mathfrak{C}_4(\mathbb{R})$, and not of sets of them, Axiom 3 is satisfied. Axiom 4 is obviously satisfied by $R_{\leq_T}$.

Since geometrical Axioms 13 to 15 about equidistance, Axioms 16 to 21 about OnL, as well as the two Five Segment Axioms 22 and 23 are valid sentences of Euclidean geometry, they must be satisfied in $\mathfrak{C}_4(\mathbb{R})$. Since they additionally are free of closure conditions, they also apply to finite subsets, and thus to F. The only geometrical axiom that involves an existential claim is Axiom 12, but it just states the existence of points $\vec{0}, \vec{a}^*, \vec{a}_1, \vec{a}_2, \vec{a}_3$, and thus is satisfied. The latter vectors are also the extension of the spatial reference frame *RefFrame*, and as such they satisfy Axioms 6 and 9, since spatially equivalent vectors all have zero distance in the three first dimensions, 7, since they are all on point-like features, and 8, since they have the required spatial configuration.

Axiom 5 about surface connectedness states that it is a partial equivalence relation. In our model, it is interpreted in terms of locational equality and therefore is satisfied. Similarly, point-like features PF are identified as a partial equivalence by instantiations of Axiom Schemas 3 and 4, which are satisfied because in our model they coincide with surfaces. Axiom Schema 5 states that point-like features have unique surface as host, which is satisfied because hosts coincide with the features in our model.

Affordance observation predicates require also partial equivalence by Axiom Schemas 9 and 10. The interpretation $R_{AffordsSeeing}$ locates $\vec{x}, \vec{y}$ on a certain ray, which is an equivalence class of vectors. Since we don't have any concrete actions observed in our model, i.e. the predicate is interpreted into the empty relation, Axiom Schemas 6 to 8 about DoX are trivially satisfied. Analogously for the locomotion affordances described in Axioms 28, 29, as well as the instantiated Axiom Schema 11, and the color predicate in Axiom 11.

Axiom 25 requires surfaces to be touched by one illuminated medium and bounded by one substance. As I have argued above, the spheres centered on $\vec{a'}_i$ and bounded by $\vec{a}_i$ are illuminated media touching the surface. Now consider the sphere centered on $\vec{0}$ and bounded by $\vec{a}_i$. Since $\vec{0}$ is not visible by Definition of $R_{AffordsSeeing}$, $\vec{a}_i$ are touched by a chain of substance contained spheres, $CSCTouched_{\neg Visible}\left(\vec{a}_i, \vec{a}_i, \vec{0}, \vec{0}\right)$. Since there are no more instances of surfaces, optic spheres and substances in the model, this makes the Axiom true. Since $R_{OnL}(\vec{a}_i, \vec{a'}_i, \vec{0})$, these touching spheres are also on a line, and thus Axiom 24 is satisfied. The substance sphere $\vec{0}\vec{a}_i$ corresponds to a body that was not reified in our model, however it satisfies Axiom 27 in either case. Furthermore, since substance and optic spheres do not overlap, Axiom 26 is satisfied.

Last but not least, verticality is also required to be a partial equivalence by Axiom 10. This is so in our model since it is interpreted into a vertical slice in space time. *q.e.d.*

Now consider the model $\mathfrak{M}_{II}$ which is just like $\mathfrak{M}_I$ except that it has one additional relation:

Definitions $R_{AffordsBreathing}(\vec{x}, \vec{y}) \leftrightarrow R_{AffordsSeeing}(\vec{x}, \vec{x}, \vec{y})$

Theorem 27 *Incompleteness:*
$\mathfrak{M}_{II}$ is a model and satisfies a sentence which is not satisfied by $\mathfrak{M}_I$.

Proof $\mathfrak{M}_{II}$ is a model since $R_{AffordsBreathing}$ satisfies Axiom Schemas 9 and 10 in analogy to $R_{AffordsSeeing}$, and nothing else was modified.

Now consider the sentence $\exists x.Air(x)$. Since m_i are illuminated media, all their elements are visibility connected. By the interpretation $R_{AffordsBreating}$, they also afford breathing. Therefore, by Definitions 58 and 59, the three illuminated media are also air media, $Air(m_i)$, and thus the sentence is satisfied by $\mathfrak{M}_{II}$. But it is not valid in $\mathfrak{M}_I$, since the breathing affordance was not observed. *q.e.d.*

8.4. List of proposed axioms

8.4.1. Axioms

8.4.2. Axiom schemas

Bibliography

[1] B. Adams and M. Raubal. A metric conceptual space algebra. In K.S. Hornsby, C. Claramunt, M. Denis, and G. Ligozat, editors, *Spatial Information Theory, 9th International Conference, COSIT 2009, Aber Wrac'h, France, September 21-25, 2009 Proceedings*, pages 51–68. Springer, Berlin, 2009.

[2] D.M. Armstrong. *Universals. An opinionated introduction.* Westview Press, Boulder, 1989.

[3] R. Arnheim. *Art and visual perception.* University of California Press, Berkeley, 50th anniversary printing edition, 2004.

[4] S. Awodey and E.H. Reck. Completeness and categoricity: 19th century axiomatics to 21st century semantics. *History and Philosophy of Logic*, 23 (77-94):1–30, 2001.

[5] L.C. Barsalou. Perceptual symbol systems. *Behavioral and Brain Sciences*, 22:577–660, 1999.

[6] L.W. Barsalou. Grounded cognition. *Annual Review of Psychology*, 59: 617–645, 2008.

[7] J. Bateman. Towards a generic foundation for spatial ontology. In A.C. Varzi and L. Vieu, editors, *Proceedings of the 3rd International Conference on Formal Ontology in Information Systems (FOIS-04)*, pages 237–248. IOS Press, Amsterdam, 2004.

[8] T. Belpaeme, L. Steels, and J. Van Looveren. The construction and acquisition of visual categories. In *Learning Robots. 6th European Workshop*, pages 1–12. Springer, Berlin, 1997.

[9] B. Bennett. Relative definability in formal ontologies. In A. Varzi and L. Vieu, editors, *Proceedings of the 3rd International Conference on Formal Ontology in Information Systems (FOIS-04)*, pages 107–118. IOS Press, 2004.

[10] B. Bennett, A.G. Cohn, P. Torrini, and S.M. Hazarika. A foundation for region-based qualitative geometry. In W. Horn, editor, *ECAI 2000 Proc.*, volume 54 of *Frontiers in artificial intelligence and applications*, pages 204–208. IOS-Press [u.a.], Amsterdam, 2000.

[11] B. Bennett, D. Mallenby, and A. Third. An ontology for grounding vague geographic terms. In C. Eschenbach and M. Gruninger, editors, *Proc. 5th*

Intern. Conf. on Formal Ontology in Information Systems, pages 280–293. IOS-Press [u.a.], Saarbrücken, 2008.

[12] J. Bentham. The theory of fictions. In C.K. Ogden, editor, *Bentham's Theory of Fictions*. Kegan Paul, Trench, Trubner, London, 1932.

[13] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web - a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American Magazine*, 2001. URL `http://www.sciam.com/article.cfm?id=the-semantic-web\&\#38;print=true`.

[14] C. Bizer, T. Heath, and T. Berners-Lee. Linked Data. The story so far. *Int. J. Semantic Web Inf. Syst.*, 5(3):1–22, 2009.

[15] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. DBpedia. A crystallization point for the web of data. *Journal of Web Semantics*, 7:154–165, 2009.

[16] A. Borgida and J. Mylopoulos. Data semantics revisited. In *Semantic Web and Databases*, volume 3372 of *Lecture notes in computer science*, pages 9–26. Springer, Berlin, 2005.

[17] S. Borgo, N. Guarino, and C. Masolo. A pointless theory of space based on strong connection and congruence. In L.C. Ailello, J. Doyle, and S.C. Shapiro, editors, *Principles of Knowledge Representation and Reasoning (KR96): Proc. of the 5th Intern. Conf.*, pages 220–229. Morgan Kaufmann Publishers, Boston, 1996.

[18] S. Borgo, N. Guarino, and C. Masolo. An ontological theory of physical objects. In L. Ironi, editor, *Qualitative reasoning. Proc. of the 11th Intern. Workshop on Qualitative Reasoning*, pages 223–231. I.A.N.-C.N.R., Pavia, 1997.

[19] R.N. Boyd. On the current status of the issue of scientific realism. *Erkenntnis*, 19:45–90, 1983.

[20] R.N. Boyd. Scientific realism (stanford encyclopedia of philosophy), 2002. URL `http://plato.stanford.edu/entries/scientific-realism/`.

[21] P.W. Bridgman. *The nature of physical theory*. Princeton University Press, New York, 1936.

[22] N. Burgess. Spatial cognition and the brain. *Ann. N. Y. Acad. Sci.*, 1124: 77–97, 2008.

[23] R. Carnap. Empiricism, semantics, and ontology. *Revue Internationale de Philosophie*, 4(1950):20–40, 1950.

[24] R. Carnap. Beobachtungssprache und theoretische sprache. *Dialectica*, 12(1958):236–248, 1958.

[25] R. Carnap. *Der logische Aufbau der Welt*. 1998.

[26] R. Casati and A.C. Varzi. *Parts and places: The structures of spatial representation*. MIT Press, Cambridge, Mass., 1999.

[27] H. Couclelis. Ontology, epistemology, teleology: Triangulating geographic information science. In G. Navratil, editor, *Research Trends in Geographic Information Science*, volume 12 of *Lecture Notes in Geoinformation and*

Cartography, pages 3–15. Springer, Berlin, 2009.

[28] H. Couclelis. Ontologies of geographic information. *International Journal of Geographical Information Science*, 24(12):1785–1809, 2010.

[29] T. Cresswell. *Place: A short introduction*. Blackwell Publishing, Oxford, UK, 2004.

[30] D. Davidson. Meaning, truth and evidence. In R. Barrett and R. Gibson, editors, *Perspectives on Quine*, pages 68–79. Blackwell, Cambridge, MA, 1990.

[31] R. Diestel. *Graph theory*, volume 173 of *Graduate texts in mathematics*. Springer, New York, NY, 2. ed edition, 2000.

[32] P. Duhem. *The aim and structure of physical theory*. Princeton University Press, Princeton, New Jersey, 1954.

[33] H.D. Ebbinghaus and J. Flum. *Finite model theory*. Springer monographs in mathematics. Springer, Berlin, 2005.

[34] U. Eco. *Zeichen. Einführung in einen Begriff und seine Geschichte*. Suhrkamp, Frankfurt am Main, 1977.

[35] U. Eco. *Kant and the platypus. Essays on language and cognition*. Vintage, London, 2000.

[36] A.S. Eddington. What is geometry? In J.J.C. Smart, editor, *Problems of space and time*, pages 163–177. The Macmillan Company, New York, 1964.

[37] E. Hyvönen et al. CultureSampo - Finnish culture on the semantic web 2.0. Thematic perspectives for the end-user. In J. Trant and D. Bearman, editors, *Museums and the Web 2009: Proceedings*, pages 98–109, Toronto, 2009. Archives and Museum Informatics.

[38] J. Euzenat and P. Shvaiko. *Ontology matching*. Springer, Heidelberg (DE), 2007.

[39] F. Flores. The equivalence of mass and energy (Stanford Encyclopedia of Philosophy), 2007. URL `http://plato.stanford.edu/entries/equivME/`.

[40] J.A. Fodor. *The language of thought*. Harvard University Press, Cambridge, MA, 1975.

[41] J.A. Fodor and Z. Pylyshyn. How direct is visual perception. *Ecological Psychology*, 9(1):139–196, 1981.

[42] A.U. Frank. Ontology for spatio-temporal databases. In M. Koubarakis et al., editor, *Spatio-temporal databases: The chorochronos approach*, pages 9–77. Springer, Berlin, 2003.

[43] A.U. Frank. Scale is introduced in spatial datasets by observation processes. In R. Devillers and M. Goodchild, editors, *Spatial Data Quality From Process to Decision (6th ISSDQ 2009)*, pages 17–29. CRC Presss, 2009.

[44] G. Frege. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, NF 100:25–50, 1892.

[45] C. Freksa and T. Barkowsky. On the relation between spatial concepts

and geographic objects. In P. Burrough and A. Frank, editors, *Geographic objects with indeterminate boundaries*, pages 109–121. Taylor Francis, London, 1996.

[46] A. Galton and R. Mizoguchi. The water falls but the waterfall does not fall: New perspectives on objects, processes and events. *Applied Ontology*, 4 (2):71–107, 2009.

[47] A. Gangemi, N. Guarino, C. Masolo, and A. Oltramari. Understanding top-level ontological distinctions. In H. Stuckenschmidt, editor, *Proc. IJCAI 2001 Workshop on Ontologies and Information Sharing*. 2001.

[48] P. Gärdenfors. Conceptual spaces as a framework for knowledge representation. *Mind and Matter*, 2(2):9–27, 2004.

[49] M.R. Genesereth and N.J. Nilsson. *Logical foundations of artifical intelligence*. Kaufmann, Los Altos, Calif., 1987.

[50] J.J. Gibson. *The ecological approach to visual perception*. Houghton Mifflin, Boston, 1979.

[51] E. v. Glasersfeld. On the concept of interpretation. *Poetics*, 12 (2/3): 207–218, 1983.

[52] E. v. Glasersfeld. How do we mean? A constructivist sketch of semantics. *Cybernetics and Human Knowing*, 6(1):9–16, 1999.

[53] M. Goodchild. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221, 2007.

[54] N. Goodman. *The Structure of Appearance*. Reidel, Dordrecht, 3rd edition, 1977.

[55] H.P. Grice. Logic and conversation. In Cole and Morgan, editors, *Speech acts*, Syntax and Semantics 3, pages 41–58. Academic Press, New York, 1975.

[56] N. Guarino. Formal ontology and information systems. In N. Guarino, editor, *Formal ontology in information systems: Proceedings of the first international conference (FOIS '98), June 6 - 8, Trento, Italy*, volume 46 of *Frontiers in artificial intelligence and applications*. IOS-Press [u.a.], Amsterdam, 1998.

[57] N. Guarino. The ontological level. In A.T. Borgida et al., editor, *Conceptual Modeling: Foundations and Applications. Essays in Honor of John Mylopoulos*, pages 52–67. Springer, Berlin, 2009.

[58] N. Guarino and C. Welty. A formal ontology of properties. In R. Dieng and O. Corby, editors, *Knowledge Acquisition, Modeling and Management: 12th International Conference, EKAW 2000*, volume 1937 of *Lecture notes in computer science*, pages 97–112. Springer, Berlin, 2000.

[59] N. Guarino and C. Welty. Identity, unity, and individuality: Towards a formal toolkit for ontological analysis. In W. Horn, editor, *ECAI 2000 Proc.*, volume 54 of *Frontiers in artificial intelligence and applications*, pages 219–223. IOS-Press, Amsterdam, 2000.

[60] G. Guizzardi. On the representation of quantities and their parts in concep-

tual modeling. In A. Galton and R. Mizoguchi, editors, *Formal Ontology in Information Systems. Proceedings of the Sixth International Conference, FOIS 2010*, pages 103–116. IOS Press, Amsterdam, 2010.

[61] C. Habel. Discreteness, finiteness, and the structure of topological spaces. In C. Eschenbach, C. Habel, and B. Smith, editors, *Topological Foundations of Cognitive Science. Papers from the Workshop at the FISI-CS, Buffalo, NY, 1994*, Graduiertenkolleg Kognitionswissenschaft (Report 37), pages 81–90. Universität Hamburg, Hamburg, 1994.

[62] H. Halpin, P.J. Hayes, J.P. McCusker, D.L. McGuinness, and H.S. Thompson. When owl:sameAs isn't the same: An analysis of identity in linked data. In P.F. Patel-Schneider et al., editor, *International Semantic Web Conference (1)*, Lecture Notes in Computer Science, pages 305–320. Springer, 2010.

[63] S. Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.

[64] J. Hawkins and S. Blakeslee. *On intelligence: How a new understanding of the brain will lead to the creation of truly intelligent machines*. Henry Holt, 2004.

[65] P.J. Hayes. The second naive physics manifesto. In J.R. Hobbs and R.C. Moore, editors, *Formal theories of the commonsense world*, Ablex series in artificial intelligence. Ablex Publ. Corp., Norwood, N.J., 1985.

[66] G. Hickok. The role of mirror neurons in speech perception and action word semantics. *Language and Cognitive Processes*, 25(6):749–776, 2010.

[67] D. Israel. A short companion to the naive physics manifesto. In J.R. Hobbs and R.C. Moore, editors, *Formal theories of the commonsense world*, Ablex series in artificial intelligence. Ablex Publ. Corp., Norwood, N.J., 1985.

[68] R. Jackendoff. What is a concept, that a person may grasp it? In *Languages of the Mind: Essays on Mental Representation*, pages 21–52. The MIT Press, Cambridge, Massachusetts, 1993.

[69] R. Jackendoff. Word meanings and what it takes to learn them. In *Languages of the mind: Essays on mental representation*, pages 53–65. The MIT Press, Cambridge, Massachusetts, 1993.

[70] K. Janowicz. The role of place for the spatial referencing of heritage data. In *The Cultural Heritage of Historic European Cities and Public Participatory GIS Workshop*. The University of York, York, UK, 2009.

[71] K. Janowicz. The role of space and time for knowledge organization on the semantic web. *Semantic Web Journal*, 1(1-2):25–32, 2010.

[72] K. Janowicz. Observation-driven geo-ontology engineering. *Transactions in GIS*, 2012. forthcoming.

[73] K. Janowicz, M. Raubal, A. Schwering, and W. Kuhn. Special issue on semantic similarity measurement and geospatial applications. *Transactions in GIS*, 12(6), 2008.

[74] M. Johnson. *The body in the mind. The bodily basis of meaning, imagination and reason*. The University of Chicago Press, Chicago, 1987.

[75] T. Jordan, M. Raubal, B. Gartrell, and M. J. Egenhofer. An affordance-based model of place in GIS. In T. Poiker and N. Chrisman, editors, *8th Int. Symposium on Spatial Data Handling (SDH'98)*, pages 98–109, Vancouver, BC, 1998. IUG.

[76] W. Kamlah and P. Lorenzen. *Logische Propädeutik. Vorschule des vernünftigen Redens.* J.B. Metzler, Stuttgart, Weimar, 3rd edition, 1996.

[77] P. Kay and L. Maffi. Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist*, 101:743–760, 1999.

[78] P. Kay and C. McDaniel. The linguistic significance of the meanings of basic color terms. *Language*, 54(3):610–646, 1978.

[79] A. Klippel. Spatial information theory meets spatial thinking - is topology the rosetta stone of spatial cognition? *Annals of the Association of American Geographers*, 2011.

[80] W. Köhler. *Gestalt Psychology. An introduction to new concepts in modern psychology.* Liveright, New York, 1992.

[81] W. Kuhn. Semantic reference systems. *Int. J. Geogr. Inf. Science*, 17 (5): 405–409, 2003.

[82] W. Kuhn. Geospatial semantics: Why, of what and how? In S. Spaccapietra and E. Zimányi, editors, *Journal on Data Semantics III*, volume 3534 of *Lecture notes in computer science*, pages 1–24. Springer, Berlin, 2005.

[83] W. Kuhn. An image-schematic account of spatial categories. In S. Winter, M. Duckham, L. Kulik, and B. Kuipers, editors, *Spatial Information Theory: 8th International Conference, COSIT 2007, Melbourne, Australia, September 19-23, 2007, Proceedings*, volume 4736, pages 152–168, Berlin, 2007. Springer.

[84] W. Kuhn. Semantic engineering. In G. Navratil, editor, *Research Trends in Geographic Information Science*, volume 12 of *Lecture Notes in Geoinformation and Cartography*, pages 63–76. Springer, Berlin, 2009.

[85] W. Kuhn. A functional ontology of observation and measurement. In *GeoSpatial Semantics. Third International Conference, GeoS 2009, Mexico City*, LNCS, pages 26–43, Berlin, 2009. Springer.

[86] W. Kuhn. Modeling vs encoding for the semantic web. *Semantic Web Journal*, 1(1-2):11–15, 2010.

[87] W. Kuhn and M. Raubal. Implementing semantic reference systems. In M. Gould, R. Laurini, and S. Coulondre, editors, *6th AGILE Conference on Geographic Information Science, Lyon, France*, pages 63–72. Presses Polytechniques et Universitaires Romandes, Lausanne, 2003.

[88] B. Kuipers. Modeling spatial knowledge. *Cognitive Science*, 2:129–153, 1978.

[89] J. Ladyman. *Understanding philosophy of science.* Rootledge, London, 2002.

[90] G. Lakoff. *Women, fire and dangerous things: What categories reveal about the mind.* Univ. of Chicago Press, Chicago, 1990.

[91] R.W. Langacker. Nouns and verbs. *Language*, 63(1):53–94, 1987.

[92] R.W. Langacker. Dynamicity, fictivity and scanning. the imaginative basis of logic and linguistic meaning. In D. Pecher and A. Zwaan, editors, *Grounding cognition. The role of perception and action in memory, language and thinking*, pages 164–197. Cambridge Univ. Press, Cambridge, 2005.

[93] S. Lawrence and E. Margolis. Concepts and cognitive science. In E. Margolis and S. Laurence, editors, *Concepts: Core readings*, pages 259–303. MIT Press, MA, 1999.

[94] S. Lehar. *The world in your head. A Gestalt view of the mechanism of conscious experience*. Lawrence Erlbaum Associates, Mahwah, London, 2003.

[95] G.W. Leibniz and S. Clarke. *The Leibniz-Clarke correspondence*. Manchester University Press, Manchester, UK, 1956.

[96] K. Lorenz. *Elemente der Sprachkritik. Eine Alternative zum Dogmatismus und Skeptizismus in der Analytischen Philosophie*. Theorie. Suhrkamp, Frankfurt, 1970.

[97] K. Lorenz and J. Mittelstraß. Die Hintergehbarkeit der Sprache. *Kant Studien*, 58(1-4):187–208, 1967.

[98] P. Lorenzen. *Einführung in die operative Logik und Mathematik*. Springer Verlag, Berlin, 1955.

[99] P. Lorenzen. Das Aktual-Unendliche in der Mathematik. *Philosophia Naturalis*, 4(1):3–11, 1957.

[100] P. Lorenzen. Methodisches Denken. *Ratio*, 7:1–23, 1965.

[101] K. Lynch. *The image of the city*. MIT Press, Cambridge MA, 1960.

[102] J. Malkevitch. Finite geometries?, 2006. URL `http://www.ams.org/featurecolumn/archive/finitegeometries.html`.

[103] G. Marchetti. A presentation of attentional semantics. *Cognitive Processsing*, 7(3):163–194, 2006.

[104] G. Marchetti. A criticism of Leonard Talmy's cognitive semantics. *self*, 2006.

[105] D. Marr. *Vision*. W.H. Freeman, San Francisco, 1982.

[106] D. Marr and H.K. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 200(1140):269–294, 1978.

[107] C. Masolo. Founding properties on measurement. In A. Galton and R. Mizoguchi, editors, *Formal Ontology in Information Systems. Proc. of the Sixth International Conference (FOIS 2010)*, pages 89–102. IOS Press, Amsterdam, 2010.

[108] C. Masolo and S. Borgo. Qualities in formal ontologies. In P. Hitzler, C. Lutz, and G. Stumme, editors, *Workshop on Foundational Aspects of Ontologies (FOnt 2005)*, volume 9/2005 of *Fachbericht Informatik*. Universität Koblenz-Landau, Koblenz, 2005.

[109] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, and A. Oltramari. Wonderweb deliverable d18: Ontology library. Trento, Italy, 2003.

[110] A. Mathes. Folksonomies - cooperative classification and communication through shared metadata, 2009. URL `http://adammathes.com/academic/computer-mediated-communication/folksonomies.html`.

[111] M. May, S. Scheider, R. Rösler, D. Schulz, and D. Hecker. Pedestrian flow prediction in extensive road networks using biased observational data. In W.G. Aref, M.F. Mokbel, and M. Schneider, editors, *GIS '08: Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, pages 471–480. ACM, New York, 2008.

[112] G.H. Mead. *Mind, self, and society.* University of Chicago Press, Chicago, 1934.

[113] M. Merleau-Ponty. *Le primat de la perception et ses consequences philosophiques.* Verdier, 1996.

[114] G.H. Merrill. Ontological realism: Methodology or misdirection? *Applied Ontology*, 5:79–108, 2010.

[115] A. Michotte, G. Thines, and G. Crabbe. Amodal completion of perceptual structures. In G Thines, A. Costall, and Butterworth G., editors, *Michotte's experimental phenomenology of perception*, pages 140–167. Lawrence Erlbaum Associates, Hillsdale, NJ, 1991.

[116] T.M. Mitchell. The need for biases in learning generalizations. Cbm-tr 5-110, Rutgers University, 1980.

[117] D.R. Montello. Scale and multiple psychologies of space. In A. U. Frank and I. Campari, editors, *Spatial information theory: A theoretical basis for GIS. Proceedings of COSIT '93*, pages 312–321, Berlin, 1993. Springer.

[118] F. Neuhaus, P. Grenon, and B. Smith. A formal theory of substances, qualities and universals. In A. Varzi and L. Vieu, editors, *Proceedings of the 3rd International Conference on Formal Ontology in Information Systems (FOIS-04)*, pages 49–59. IOS Press, 2004.

[119] A. Newell and H.A. Simon. Computer science as empirical inquiry: Symbols and search. *Commun. ACM*, 19(3):113–126, 1976.

[120] N.J. Nilsson. The physical symbol system hypothesis: Status and prospects. In *50 Years of Artificial Intelligence*, pages 9–17, 2006.

[121] H. Noonan. Identity (stanford encyclopedia of philosophy), 2009. URL `http://plato.stanford.edu/entries/identity/`.

[122] C.K. Ogden and I.A. Richards. *The meaning of meaning.* Harcourt, Brace & World, New York, 1923.

[123] J. Ortmann and W. Kuhn. Affordances as qualities. In A. Galton and R. Mizoguchi, editors, *Formal Ontology in Information Systems. Proc. of the 6th Int. Conf.*, LNCS, pages 117–130. IOS Press, Amsterdam, 2010.

[124] R.R. Oudejans, C.F. Michaels, B. van Dort, and E.J.P. Frissen. To cross or not to cross: The effect of locomotion on street-crossing behavior. *Ecological Psychology*, 8(3):259–267, 1996.

[125] S.E. Palmer. Fundamental aspects of cognitive representation. In E. Rosch and B.B. Lloyd, editors, *Cognition and Categorization*, pages 259–303. Erlbaum, Berkeley?, 1978.

[126] J. Piaget. *Genetic epistemology. Woodbridge lecture no. 8.* Columbia University Press, New York and London, 1st edition, 1970.

[127] C. Pincock. Carnap's logical structure of the world. *Philosophy Compass*, 4:951–961, 2009.

[128] H. Poincaré. *Science and hypothesis.* Dover Publ., N.Y., 1952.

[129] J. Possin. Affordance-basierte Interpretation von Openstreetmap-Daten für Navigationsanwendungen. Diplomarbeit, University of Münster, Münster, Germany, 2011.

[130] F. Probst. *Semantic reference systems for observations and measurements.* PhD thesis, University of Münster, 2007.

[131] F. Probst. Observations, measurements and semantic reference spaces. *Applied Ontology*, 3(1-2):63–89, 2008.

[132] F. Probst and M. Espeter. Spatial dimensionality as classification criterion for qualities. In B. Bennett and C. Fellbaum, editors, *Formal Ontology in Information Systems: Proceedings of the Fourth International Conference: FOIS 2006*, volume 150 of *Frontiers in artificial intelligence and applications*, pages 77–88. IOS Press, 2006.

[133] H. Putnam. The meaning of 'meaning'. In H. Putnam, editor, *Mind, language, and reality*, volume 2 of *Philosophical Papers*, pages 215–271. Cambridge University Press, Cambridge, 1979.

[134] H. Putnam. *Reason, truth and history.* Cambridge University Press, Cambridge, 1981.

[135] Z.W. Pylyshyn. *Things and places. How the mind connects with the world.* The MIT Press, Cambridge, Massachusetts, 2007.

[136] W.V.O. Quine. Two dogmas of empiricism. *The Philosophical Review*, 60: 20–43, 1951.

[137] W.V.O. Quine. *The roots of reference.* Open Court Publishing, La Salle, Illinois, 1974.

[138] W.V.O. Quine. On what there is. In *From a logical point of view. 9 logico-philosophical essays,2nd edition.* Harvard University Press, Cambridge, Massachusetts, 2nd edition, 1980.

[139] W.V.O. Quine. Identity, ostension and hypostasis. In *From a logical point of view. 9 logico-philosophical essays*, pages 65–79. Harvard University Press, Cambridge, Massachusetts, 2nd edition, 1980.

[140] W.V.O. Quine. Empirical content. In W.V.O. Quine, editor, *Theories and Things*, pages 24–30. Harvad University Press, Cambridge, Massachusetts, 1981.

[141] W.V.O. Quine. Things and their place in theories. In W.V.O. Quine, editor, *Theories and Things*, pages 1–23. Harvad University Press, Cambridge, Massachusetts, 1981.

[142] W.V.O. Quine. *Pursuit of truth*. Harvard University Press, Cambridge, Massachusetts, 1990.

[143] W.V.O. Quine. *Word and object*. MIT Press, Cambridge, Mass., 24. pr. edition, 2001.

[144] M.J. Reddy. The conduit metaphor: A case of frame conflict in our language about language. In Andrew Ortony, editor, *Metaphor and Thought*, pages 164–201. Cambridge University Press, Cambridge, 1993.

[145] G. Rizzolatti and C. Sinigaglia. *Mirrors in the brain: How our minds share actions and emotions*. Oxford University Press, Oxford, 2008.

[146] L.A. Rodegerdts. Signalized intersections: Informational guide., 2009. URL `http://www.tfhrc.gov/safety/pubs/04091/`.

[147] B. Russell. The philosophy of logical atomism. In D. Pears, editor, *The philosophy of logical atomism*, pages 35–156. Open Court, La Salle, Illinois, 1985.

[148] J.T. Sanders. An ontology of affordances. *Ecological Psychology*, 9 (1): 97–112, 1997.

[149] E.E. Savellos. On defining identity. *Notre Dame Journal of Formal Logic*, 31(3):476–484, 1990.

[150] A. Scarantino. Affordances explained. *Philosophy of Science*, 70:949–961, 2003.

[151] S. Scheider. The case for grounding databases. In K. Janowicz, M. Raubal, and S. Levashkin, editors, *GeoSpatial Semantics. Third International Conference, GeoS 2009, Mexico City*, LNCS, pages 44–62. Springer, Berlin, 2009.

[152] S. Scheider and K. Janowicz. Places as media of containment. In R. Purves and R. Weibel, editors, *Extended Abstracts Volume, GIScience 2010*. University Zürich-Irchel, Zürich, 2010.

[153] S. Scheider and W. Kuhn. Road networks and their incomplete representation by network data models. In T.J. Cova, H.J. Miller, K. Beard, A.U. Frank, and M.F. Goodchild, editors, *Geographic Information Science: 5th International Conference, GIScience 2008, Park City, UT, USA, September 23 - 26, 2008, Proceedings*, pages 290–307. Springer, Berlin, 2008.

[154] S. Scheider and W. Kuhn. Affordance-based categorization of road network data using a grounded theory of channel networks. *International Journal of Geographical Information Science*, 24(8):1249 – 1267, 2010.

[155] S. Scheider and W. Kuhn. Finite relativist geometry grounded in perceptual operations. In M. et al. Egenhofer, editor, *Spatial Information Theory: 10th International Conference, COSIT 2011,*, LNCS 6899, pages 304–327. Springer, Berlin, 2011.

[156] S. Scheider and J. Possin. Affordance-based individuation of junctions in open street map. *Journal of Spatial Information Science*, (4), 2012.

[157] S. Scheider and D. Schulz. Specifying essential features of street networks. In S. Winter, M. Duckham, L. Kulik, and B. Kuipers, editors, *Spatial In-*

formation Theory: 8th International Conference, COSIT 2007, Melbourne, Australia, September 19-23, 2007, Proceedings, volume 4736/2007, pages 169–185. Springer, Berlin, 2007.

[158] S. Scheider, K. Janowicz, and W. Kuhn. Grounding geographic categories in the meaningful environment. In K.S. Hornsby, C. Claramunt, Denis M., and G. Ligozat, editors, *Spatial Information Theory, 9th International Conference, COSIT 2009, Aber Wrac'h, France, September 21-25, 2009 Proceedings*, pages 69–87. Springer, Berlin, 2009.

[159] S. Scheider, F. Probst, and K. Janowicz. Constructing bodies and their qualities from observations. In A. Galton and R. Mizoguchi, editors, *Formal Ontology in Information Systems. Proc. of the Sixth International Conference (FOIS 2010)*, pages 131–144. IOS Press, Amsterdam, 2010.

[160] S. Scheider, C. Kessler, J. Ortmann, A. Devaraju, J. Trame, T. Kauppinen, and W. Kuhn. Semantic referencing of geosensor data and volunteered geographic information. In N. Ashish and A.P. Sheth, editors, *Geospatial Semantics and the Semantic Web. Foundations, Algorithms, and Applications*, volume 12 of *Semantic Web and Beyond*, pages 27–59. Springer, Berlin, 2011.

[161] B.J. Scholl. Objects and attention: The state of the art. *Cognition*, 80: 1–46, 2001.

[162] W. Schwabhäuser, W. Szmielew, and A. Tarski. *Metamathematische Methoden in der Geometrie, Teil I: Ein axiomatischer Aufbau der euklidischen Geometrie*. Springer Verlag, Berlin, 1983.

[163] J.R. Searle. *Speech Acts*. Cambridge University Press, Cambridge, UK, 1969.

[164] J.R. Searle. Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3):417–457, 1980.

[165] J.R. Searle. Rationality and realism: What is at stake. *Daedalus*, 122(4): 55–84, 1992.

[166] J.R. Searle. Social ontology: Some basic principles. *Anthropological Theory*, 6:12–29, 2006.

[167] J. Seibt. Beyond endurance and perdurance: Recurrent dynamics. In Ch. Kanzian, editor, *Persistence*, pages 133–165. Ontos Verlag, Frankfurt, 2007.

[168] R.E. Shaw, M.T. Turvey, and W. Mace. Ecological psychology: The consequence of a commitment to realism. In W.B. Weimer and D.S. Palermo, editors, *Cognition and the Symbolic Processes*, volume 2, pages 159–226. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1982.

[169] A.P. Sheth. Changing focus on interoperability in information systems: From system, syntax, structure to semantics. In R. Fegeas M.F. Goodchild, M.J. Egenhofer and C.A. Kottman, editors, *Interoperating Geographic Information Systems*, pages 5–30. Kluwer, Norwell, MA, USA, 1999.

[170] A.P. Sheth, C. Henson, and S.S. Sahoo. Semantic sensor web. *Internet Computing, IEEE*, 12(4):78–83, 2008.

[171] C. Shirky. Ontology is overrated, 2009. URL `http://www.shirky.com/writings/ontology\_overrated.html`.

[172] A. Sloman. Interactions between philosophy and artificial intelligence: The role of intuition and non-logical reasoning in intelligence. *Artificial Intelligence*, 2:209–225, 1971.

[173] B. Smith. Beyond concepts: Ontology as reality representation. In A. Varzi and L. Vieu, editors, *Proceedings of the 3rd International Conference on Formal Ontology in Information Systems (FOIS-04)*, pages 73–84. IOS Press, 2004.

[174] E. Sober. What is the problem of simplicity? In A. Zellner, H. Keuzenkamp, and M. McAleer, editors, *Simplicity, Inference, and Modelling*, pages 13–32. Cambridge University Press, Cambridge, Massachusetts, 2002.

[175] C. Stasch, K. Janowicz, A. Bröring, I. Reis, and W. Kuhn. A stimulus-centric algebraic approach to sensors and observations. In *3rd International Conference on Geosensor Networks*, volume 5659 of *LNCS*, pages 169–179. Springer, 2009.

[176] L. Steels. A self-organizing spatial vocabulary. *Artificial Life Journal*, 2(3): 319–332, 1996.

[177] L. Steels. The symbol grounding problem has been solved. So what's next? In M. de Vega, editor, *Symbols and Embodiment: Debates on Meaning and Cognition*, chapter 12. Oxford University Press, Oxford, 2008.

[178] T. A. Stoffregen. Affordances as properties of the animal-environment system. *Ecological Psychology*, 15 (2):115–134, 2003.

[179] P.F. Strawson. *Individuals. An essay on descriptive metaphysics.* Routledge, London, 1959.

[180] P. Suppes. Is visual space Euclidian? *Synthese*, 35:397–421, 1977.

[181] P. Suppes. Finitism in geometry. *Erkenntnis*, 54:133–144, 2001.

[182] P. Suppes and J.L. Zinnes. Basic measurement theory. In R. Luce, editor, *Chapters 1-8*, volume Vol. 1 of *Handbook of mathematical psychology*, pages 1–76. Wiley, New York, 1967.

[183] L. Talmy. The relation of grammar to cognition. In L. Talmy, editor, *Toward a Cognitive Semantics - Vol. I*, pages 21–96. The MIT Press, Cambridge, Mass., 2000.

[184] L. Talmy. Fictive motion in language and "ception". In L. Talmy, editor, *Toward a Cognitive Semantics - Vol. I*, pages 99–175. The MIT Press, Cambridge, Mass., 2000.

[185] L. Talmy. The windowing of attention in language. In L. Talmy, editor, *Toward a Cognitive Semantics - Vol. I*, pages 257–309. The MIT Press, Cambridge, Mass., 2000.

[186] A. Tarski. What is elementary geometry. In P. Suppes L. Henkin and A. Tarski, editors, *The axiomatic method. With special reference to geometry and physics*, pages 16–29. North-Holland Publishing, Amsterdam, 1959.

[187] A. Tarski. Tarski's system of geometry. *The Bulletin of Symbolic Logic*, 5

(2):175–214, 1999.

[188] M. Tomasello. *The cultural origins of human cognition*. Harvard University Press, Cambridge, MA, 1999.

[189] P. Turner. Affordance as context. *Interacting with Computers*, 17(6):787–800, 2005.

[190] M.T. Turvey. Affordances and prospective control: An outline of the ontology. *Ecological Psychology*, 4 (3):173–187, 1992.

[191] B. Tversky. Structures of mental spaces: How people think about space. *Environment and Behaviour*, 35(1):66–80, 2003.

[192] H. v. Foerster. Notes on an epistemology of living things. Bcl report 9.3, University of Illinois, 1972.

[193] H. v. Foerster. Objects: Tokens for (eigen-)behaviors. *ASC Cybernetics Forum*, 8(3/4):91–96, 1976.

[194] E. v. Glasersfeld. An attentional model for the conceptual construction of units and number. *Journal for Research in Mathematics Education*, 12(2): 83–94, 1981.

[195] E. v. Glasersfeld. *Radical Constructivism: A way of knowing and learning*. The Falmer Press, London, 1995.

[196] J. v. Uexküll. *Streifzüge durch die Umwelten von Tieren und Menschen*. Rowohlt, Hamburg, 1956.

[197] H. Vaihinger. *Die Philosophie des Als Ob. System der theoretischen, praktischen und religiösen Fiktionen der Menschheit auf Grund eines idealistischen Positivismus*. VDM Verlag Dr. Müller, Saarbrücken, 2007.

[198] R. VanRullen and C. Koch. Is perception discrete or continuous? *Trends in Cognitive Sciences*, 7(5):207–213, 2003.

[199] P. Vogt. Language evolution and robotics: Issues on symbol grounding and language acquisition. In A. Loula, R. Gudwin, and J. Queiroz, editors, *Artificial Cognition Systems*, pages 176–209. Idea Group, Hershey, PA., 2006.

[200] H. Wache, T. Vögele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and S. Hübner. Ontology-based integration of information: A survey of existing approaches. In H. Stuckenschmidt, editor, *Proc. IJCAI 2001 Workshop on Ontologies and Information Sharing*, pages 108–117. 2001.

[201] W.H. Warren. Perceiving affordances: Visual guidance of stair climbing. *Journal of Experimental Psychology: Human Perception and Performance*, 10(5):683–703, 1984.

[202] D. Weinberger. The machine that would predict the future. *Scientific American*, 305(6):323–340, 2011.

[203] Wikipedia. Matter, 2009. URL `http://en.wikipedia.org/wiki/Matter\#Discussion\_and\_background`.

[204] T. Williamson. *Identity and discrimination*. Blackwell, Oxford, 1990.

[205] S. Winter. Modeling costs of turns in route planning. *GeoInformatica*, 6 (4):345–360, 2002.

[206] L. Wittgenstein. *Tractatus logico-philosophicus*. Kegan Paul Ltd., London, 1922.

[207] L. Wittgenstein. *Philosophische Untersuchungen*. Suhrkamp, Frankfurt a. M., 2003.

[208] S. Wrobel. Inductive logic programming. In Gerhard Brewka, editor, *Principles of knowledge representation*, pages 153–189. CSLI Publ. [u.a.], Stanford, Calif., 1996.