Synthese Library 359

Frank Zenker
Peter Gärdenfors *Editors*

Applications of Conceptual Spaces

The Case for Geometric Knowledge Representation

Springer

Synthese Library

Studies in Epistemology, Logic, Methodology, and Philosophy of Science

Volume 359

More information about this series at http://www.springer.com/series/6607

Frank Zenker • Peter Gärdenfors
Editors

Applications of Conceptual Spaces

The Case for Geometric Knowledge Representation

Springer

Editors
Frank Zenker
Department of Philosophy and Cognitive Science
Lund University
Lund, Sweden

Peter Gärdenfors
Department of Philosophy and Cognitive Science
Lund University
Lund, Sweden

Synthese Library
ISBN 978-3-319-15020-8 ISBN 978-3-319-15021-5 (eBook)
DOI 10.1007/978-3-319-15021-5

Library of Congress Control Number: 2015937356

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Foreword

All chapters of this volume arise from the international conference *Conceptual Spaces at Wor*k, held on 24–26 May 2012 at Lund University, Sweden. We gratefully acknowledge the generous sponsorship of *The Swedish Research Council* and *The Royal Swedish Academy of Letters, History and Antiquities*.

As organizers, we regret that some contributions to this meeting do not appear in the present volume. As editors, we are indebted to a number of anonymous reviewers as well as to Christi Lue and Joos Walbeek at Springer's Dordrecht office for their assistance in the publication process.

Lund, Sweden
December 2014

Frank Zenker
Peter Gärdenfors

Contents

Part I Introduction

1 Editors' Introduction: Conceptual Spaces at Work 3
Peter Gärdenfors and Frank Zenker

Part II Semantic Spaces

2 From Conceptual Spaces to Predicates 17
Jean-Louis Dessalles

3 Conceptual Spaces at Work in Sensory Cognition: Domains, Dimensions and Distances 33
Carita Paradis

4 Conceptual Spaces, Features, and Word Meanings: The Case of Dutch Shirts 57
Joost Zwarts

5 Meaning Negotiation 79
Massimo Warglien and Peter Gärdenfors

Part III Computing Meanings

6 How to Talk to Each Other via Computers: Semantic Interoperability as Conceptual Imitation 97
Simon Scheider and Werner Kuhn

7 Conceptual Spaces and Computing with Words 123
Janet Aisbett, John T. Rickard, and Greg Gibbon

8 Self-Organisation of Conceptual Spaces from Quality Dimensions ... 141
Paul Vogt

9 Logical, Ontological and Cognitive Aspects of Object Types and Cross-World Identity with Applications to the Theory of Conceptual Spaces 165
Giancarlo Guizzardi

10 A Cognitive Architecture for Music Perception Exploiting Conceptual Spaces 187
Antonio Chella

Part IV Philosophical Perspectives

11 Conceptual Spaces as Philosophers' Tools 207
Lieven Decock and Igor Douven

12 Specification of the Unified Conceptual Space, for Purposes of Empirical Investigation 223
Joel Parthemore

13 A Perspectivist Approach to Conceptual Spaces 245
Mauri Kaipainen and Antti Hautamäki

14 Communication, Rationality, and Conceptual Changes in Scientific Theories 259
Frank Zenker and Peter Gärdenfors

Part I
Introduction

Chapter 1
Editors' Introduction: Conceptual Spaces at Work

Peter Gärdenfors and Frank Zenker

Abstract This introductory chapter provides a non-technical presentation of conceptual spaces as a representational framework for modeling different kinds of similarity relations in various cognitive domains. Moreover, we briefly summarize each chapter in this volume.

1.1 Conceptual Spaces

1.1.1 Three Kinds of Cognitive Representations

Humans are extremely efficient at learning new concepts. After having been presented with only a couple of examples, we are able to abstract the general content of a new concept. A central problem for cognitive science is how this learning process and the underlying representations should be modeled. There have been two dominating approaches to these problems. The symbolic approach starts from the assumption that cognitive systems can be described as Turing machines. On this view, cognition is seen as essentially being computation involving symbol manipulation. The second approach is associationism, where associations between different kinds of information elements carry the main burden of representation. Connectionism is a special case of associationism that models associations using artificial neuron networks.

There are aspects of cognitive phenomena, however, for which neither symbolic representation nor associationism seems to offer appropriate modeling tools. In particular, mechanisms of concept learning cannot be given a satisfactory treatment in any of these representational forms. Concept learning is closely tied to the notion of *similarity*, which has turned out to be problematic to model in the symbolic and associationist approaches.

P. Gärdenfors (✉) • F. Zenker
Department of Philosophy and Cognitive Science, Lund University,
Box 192, 221 00 Lund, Sweden
e-mail: peter.gardenfors@lucs.lu.se; frank.zenker@fil.lu.se

F. Zenker, P. Gärdenfors (eds.), *Applications of Conceptual Spaces*,
Synthese Library 359, DOI 10.1007/978-3-319-15021-5_1

A third form of representing information that employs *geometric* structures rather than symbols or associations had been presented in the book *Conceptual Spaces: The Geometry of Thought* (Gärdenfors 2000). Information is represented by points, vectors and regions in dimensional spaces. On the basis of these structures, similarity relations can be modeled in a natural way in terms of distances in a space.

The geometric approach to knowledge representation having received more attention over the last 15 years, this book aims at presenting some of its areas of application and development.

1.1.2 Conceptual Spaces as a Representational Framework

A conceptual space consists of a number of *quality dimensions*. Examples of such dimensions are: color, pitch, temperature, weight, and the three ordinary spatial dimensions. These dimensions are closely connected to what is produced by our sensory receptors (Schiffman 1982). However, there are also quality dimensions of an abstract, non-sensory character. In Gärdenfors (2007), for instance, the analysis has been extended to functional and action categories, and to event categories in Gärdenfors and Warglien (2012), all of which are treated in Gärdenfors (2014).

The primary function of quality dimensions is to represent various "qualities" of objects in different *domains*. The notion of a dimension should be understood literally. It is assumed that each of the quality dimensions is endowed with certain *topological* or *geometric* structures. Some quality dimensions are *integral* in the sense that one cannot fully describe an object by assign to it a value on one dimension without also giving it a value on others. For example, an object cannot be given a hue without also giving it a brightness value. Or the pitch of a sound always goes along with a particular loudness. Dimensions that are not integral are said to be *separable*, as for example the size and hue dimensions. Using this distinction, the notion of a *domain* can now be defined as a set of integral dimensions that are separable from all other dimensions. For an exact definition, see Zenker and Gärdenfors (Chap. 14, this volume).

Conceptual spaces are particularly suited to represent different kinds of similarity relations: the closer two objects are located in a conceptual space, the more similar they are; "green," for instance, is closer to "blue" than to "red." If dimensions are assumed to have a metric, moreover, one can talk about *distances* in the conceptual space such that distances represent degrees of similarity between the objects represented in the space.

It is important to introduce a distinction between a *psychological* and a *scientific* interpretation of quality dimensions. The psychological interpretation generally concerns how humans structure their perceptions. Vogt's Chap. 7 in this volume provides one model of how conceptual spaces can evolve from sensory quality dimensions. It is further assumed that these quality dimensions form the basis of word meanings, at least of the basic words that children learn first. A psychologically interesting example of a domain remains *color perception*, to which several

authors in this volume refer. The scientific interpretation, in contrast, deals with how different dimensions are presented within a scientific theory, how they can give rise to empirical theories, and how to model diachronic changes as science develops (see Gärdenfors and Zenker 2013; Zenker and Gärdenfors 2013; Chap. 14, this volume).

1.1.3 Properties and Concepts

Among others, the theory of conceptual spaces has been used to provide a definition of what constitutes a *natural property*. With the following criterion (Gärdenfors 1990, 1992, 2000, 2014), the geometric characteristics of the quality dimensions are utilized to introduce a spatial structure for properties:

Criterion P: A natural property is a convex region in some domain.

A set is said to be *convex* if, for all points x and y in the set, all points between x and y are also in the set. Criterion P presumes, of course, that the notion of betweenness is meaningful for the relevant quality dimensions. Being a weak assumption, this demands rather little of the underlying geometric structure of a domain.

Most properties that natural languages express by simple words seem to be natural properties in the sense specified here (Gärdenfors 2014). For instance, all *color terms* in natural languages express natural properties with respect to the psychological representation of the three color dimensions. It is well-known that different languages carve up the color circle in different ways (Berlin and Kay 1969), but all such carvings seemingly occur in terms of convex sets (Jäger 2010).

Properties, as defined by criterion P, form a special case of *concepts*. More specifically, a property is based on a *single* domain, while a concept may be based on *several* domains.

The distinction between properties and concepts has been obliterated in both the symbolic and connectionist representations. In particular, both properties and concepts are represented by *predicates* in first-order languages. The predicates of a first-order language, however, correspond to several different grammatical categories in a natural language, the most important of which are adjectives, nouns and verbs. As a development of the notions in Gärdenfors (2000), Dessalles argues in his chapter that one should distinguish between concepts that are dependent on an underlying conceptual space, and predicates which are constructed "on the fly" in a particular context.

The main semantic difference between adjectives and nouns, on the one hand, is that adjectives such as "red," "tall," and "round" normally refer to a single domain and thus represent properties, while nouns like "dog," "apple," and "town" normally contain information about several domains, and thus represent concepts. Verbs, on the other hand, obtain their meaning from their role in *events*, expressing either the action being performed ("manner verbs") or the outcome of an action ("result verbs") (Warglien et al. 2012; Gärdenfors 2014). In the event model proposed by

Gärdenfors and Warglien (2012), for instance, actions are modelled as force vectors, or patterns thereof, and results as vectors in property domains. Another example is provided in Chella et al. (2001b), who report a conceptual space describing robot actions.

Concepts are not just bundles of properties. The proposed representation for a concept also includes an account of the *correlations* between the regions of the different domains that are associated with a concept. In the "apple" concept, for instance, a very strong (positive) correlation obtains between the sweetness in the taste domain and the sugar content in the nutrition domain, while a weaker correlation holds between the color red and a sweet taste.

These considerations motivate the following definition[1]:

Criterion C: A *concept* is represented as a set of convex regions in a number of domains together with information about how the regions in different domains are correlated.

The kind of representation proposed in Criterion C is prima facie similar to *frames* (Barsalou 1992) with slots for different *features* that have been very popular within cognitive science, linguistics, and computer science. The criterion is richer, however, since a representation based on conceptual spaces allows one to describe the structure of concepts such that objects are more or less *central* representatives of a concept. Conceptual spaces thus amount to more than a combination of extant ideas from frame theory and prototype theory, since the geometry of the domains yields predictions that are not possible in either. (For a comparison of frames with conceptual spaces, see Zenker (2014)). In his contribution to this volume, Zwarts demonstrates how existing feature analyses for particular domains can be used to construct conceptual spaces in which notions like convexity can be systematically studied.

The notion of a concept defined here displays several similarities with the *image schemas* studied in cognitive linguistics, among others by Lakoff (1987) and Langacker (1987). Although image schematic representations are often pictorial, they generally fail to specify the geometric structures of the underlying domains.

1.1.4 Prototypes and Conceptual Spaces

Criterion P receives independent support from the *prototype theory* of categorization developed by Rosch and her collaborators (e.g., Rosch 1975, 1978; Mervis and Rosch 1981; Lakoff 1987). The main idea in this theory is that within a category of objects, like those instantiating a property or a concept, certain members are judged to be more representative of the category than others. Robins, for instance,

[1] A slightly more complex definition, also involving the *context* where a concept is used, is proposed in Gärdenfors (2000, ch. 4).

are taken to be more representative of the category "bird" than ravens, penguins and emus; and desk chairs are more typical instances of the category "chair" than rocking chairs, deck-chairs, and beanbag chairs. The most representative ones of a category are called its prototypical members. As Decock and Douven suggest in their chapter, some kinds may not have unique prototypes, which leads them to adapt Voronoi diagrams (introduced below) in order to deal with such vague concepts.

When natural properties are defined as convex regions of a conceptual space, prototype effects are indeed to be expected. In a convex region one can describe positions as being more or less central. In particular, if the space has a metric, one can calculate the "center of gravity" of a region.

One may also argue in the opposite direction and show that, if prototype theory is adopted, then the representation of properties as convex regions is to be expected, at least in metric spaces. To see this, assume that some quality dimensions of a conceptual space S are given (e.g., the dimensions of color space), and that we want to partition S into a number of categories (e.g., color categories). Starting from a set of prototypes $p_1, \ldots, p_n$ of such categories (such as the focal colors), these prototypes should thus be the central points in the categories they represent. On the additional assumption that S is a metric space, information about prototypes can now be used to generate a categorization. If we assume S to be equipped with the Euclidean metric, for instance, then—for every point p in S—we can measure the distance from p to each p_i. Moreover, by stipulating p to belong to the same category as the *closest* prototype p_i, such measurements generate a so-called *Voronoi tessellation* of the conceptual space, which is illustrated for the case of a plane in Fig. 1.1.

A crucial property of the Voronoi partitioning of a conceptual space is that a tessellation based on a Euclidean metric always results in a partitioning of the space into *convex* regions (see Okabe et al. 1992).

Thus, assuming that a Euclidean metric is defined on the subspace that is subject to categorization, a set of prototypes will on this method generate a unique partitioning of the subspace into convex regions. The upshot is that an intimate link obtains between prototype theory and Criterion P. Furthermore, the metric is

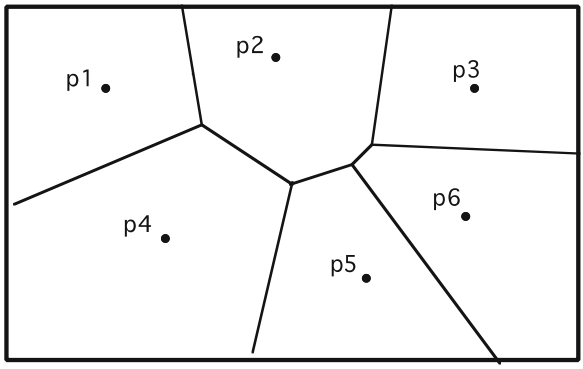

Fig. 1.1 Voronoi tessellation of the plane into convex sets, where *p1* to *p6* denote prototypes of a category, and the *lines* indicate category-boundaries

an obvious candidate for a measure of similarity between different objects. In this way, the Voronoi tessellation provides a constructive geometric answer to how a similarity measure, together with a set of prototypes, determines a set of categories.

This concludes our brief summary of the theory of conceptual spaces. The developments since the publication of Gärdenfors (2000) have shown that geometric models of knowledge representation add significantly to our explanatory capacities, particularly to understanding cognitive processes connected with learning, concept formation, (non-monotonic) reasoning, and semantics. Such conceptual representations, moreover, prove helpful in understanding how we *communicate* about concepts and thus reach semantic agreement (Gärdenfors 2014; Warglien and Gärdenfors, Chap. 5, this volume). Moreover, as Scheider and Kuhn's chapter shows, for instance, such representations allow distinguishing concepts underlying data and digital information, and thus can assist people in sharing information, particularly when interoperating across computers.

1.2 Overview of Chapters

To provide readers with further orientation, we now provide an overview of the contributions to this volume which—admittedly somewhat artificially—are grouped into the following overlapping categories: *Semantic spaces*, *computing meanings* and *philosophical perspectives*. These chapters either present successful applications of conceptual spaces theory in various research-areas, contrastive work, or research that seeks to further develop conceptual spaces.

1.2.1 Semantic Spaces

The theory of conceptual spaces builds on a cognitivist view of semantics. This contrasts with both extensional and intensional realistic semantics that include the referent of a linguistic expression as a meaning constituent. Conceptual knowledge is rather viewed as mental representations, modeled in conceptual spaces, which normally does without postulating a mental language ("Mentalese"). Therefore, such representations are not viewed as parts of a symbolic system with a syntactic or logical structure. They are instead treated as spatial structures that can be analyzed into their constitutive dimensions and properties, representing the semantic knowledge of an agent (Gärdenfors 2000, 2014).

Jean-Louis Dessalles argues that, in addition to the structure provided by conceptual spaces, we also need the operation of *contrast*. A red face, for instance, is called "red" because it contrasts with other possible face colors, rather than being red in the prototypical sense of "red". Contrast, as it develops during a conversation, is an essential operation that converts perceptions into the predicates expressed in communication. Thus *concepts*, which belong to a conceptual space, should be

kept separate from *predicates* that are formed by a contrast operation during a conversation. According to Dessalles, predicates are dynamic representations that lack a context-independent existence. Among others, he claims that the distinction between concepts (lexical meanings) and predicates serves to avoid many of the theoretical difficulties of traditional semantics.

Carita Paradis deals with how human experiences of sensory stimuli such as vision, smell, taste, and touch are rendered in natural language. Next to being part of embodied cognition, sensuous cognition also deserves attention as an important source of semantic structure. To this end, she studies data from terminological schemas for wine descriptions in wine reviews with a focus on the types of property expressions, e.g., *soft*, *sharp*, *sweet*, and *dry*, as well as object descriptors, e.g., *blueberry*, *apple* and *honey*, that are used for the different sensory experiences, and on the cross-sensory uses of these property expressions and object descriptors. In contrast to the standard view, she argues that, for instance, *sharp* in *sharp smell* does not evoke a notion of touch. When being instantiated in the sensory domain of smell, *sharp* spans closely related sensory domains. These cross-sensory uses are viewed as symptoms of synesthesia in wine-tasting and the workings of human sensory cognition.

Joost Zwarts explores how to construct a conceptual space for word meanings on the basis of a system of features, his empirical basis being Dutch words for different kinds of shirts. He considers features of shirts along eight dimensions (e.g., shape, length, collar, fabric), where a few values on each dimension corresponds to these features. Based on the Hamming distance, he then defines a metric by which to compare different kinds of shirts. Investigating this space, he shows how the different kinds are related to one another, for instance, how close to another they are in the defined conceptual space. An empirical problem is that the shirts in his empirical basis cover only a small part of the logically possible combinations of features. Although clusters of shirt types can be detected, no simple mapping between the Dutch words and locations in the space is forthcoming. Therefore, the meanings of the different words, for instance, cannot be defined as a conjunction of a set of features. The sets of shirts that correspond to a given word do nonetheless cohere in terms of the underlying feature space.

Massimo Warglien and *Peter Gärdenfors* present a model of how meanings are negotiated, that is, the process by which agents, who each start from a different preferred conceptual representation, may converge to reach an agreement through some communication medium. They show that a model based on conceptual spaces inherits important structural elements from game-theoretic models of bargaining, particularly so if agents share overlapping negotiation regions, and thus emphasize a parallel to the Nash solution in cooperative game-theory. Should agents have disjoint solution regions, moreover, processes such as changes in the salience of dimensions, dimensional projections, and metaphorical space-transformations may be helpful in finding mutually acceptable solutions. Importantly, these processes are not motivated by normative or rationality considerations, but presented as argumentation tools used in actual situations of conceptual disagreement.

1.2.2 Computing Meanings

The second part of the volume brings up several computational issues in relation to applications of conceptual spaces. Unlike most semantic theories within linguistics, the dimensional structure of conceptual spaces is useful in generating productive computational implementations.

Paul Vogt uses an agent-based simulation model to investigate how conceptual spaces can evolve from a set of quality dimensions. The model builds on the Talking Heads experiment developed by Steels and others (Steels et al. 2002), where a set of agents play a large number of guessing games that incur rewards for successful communication. Vogt's model involves a 4-dimensional conceptual space that can be broken up into 3-dimensional color and 1-dimensional shape. His results show that, upon a generational turnover occurring in the set of speakers, communication evolves towards an optimal system such that the grammar represents rules for combining color with shape. The simulations also show the evolution of the conceptual spaces which underlie communication to be driven by five crucial factors: environment, embodiment, cognition, self-organization and cultural transmission.

Simon Scheider and *Werner Kuhn* apply conceptual spaces by way of contributing to semantic technology, and seek to improve on extant models of information sharing in human-machine-human conversations. They pursue a pragmatic approach to semantic interoperability, where conceptual spaces ground (i.e., provide the constructive basis for) *learning*—which here is an attempt at imitating conceptual content that had originated under a different perspective. Accordingly, they argue, full semantic interoperability "should be defined as [full] correspondence of conceptual perspectives of communicating agents." The range of perspectival correspondence reaches from equivalent concepts, over (partially) comparable, to overlapping ones. For this purpose, empirical measurement-points are projected into a space whose convex regions are identified as concepts that are relevant for learning, and are illustrated through the reconstruction of cultural, scientific, and administrative land cover categories.

Janet Aisbett, *John Rickard* and *Greg Gibbon* explore both synergies and gaps in research on conceptual spaces. They particularly work on fuzzy sets, also known as *computing with words* (CWW), which similarly attempts to address aspects of human cognition such as judgment and intuition. Outlining formal methods developed in CWW for modelling and manipulating constructs whenever membership values are imprecise, the authors describe and problematize a specific formalism of conceptual spaces based on fuzzy sets. This formalism is compared to alternative methods for aggregating property membership into concept membership. Their problem solution is a model where all constructs are fuzzy sets on a plane, and similarity of two constructs is an inverse function of the average separation between the membership functions.

Giancarlo Guizzardi presents a system of modal logic that distinguishes sortals from general property types, and can thus capture the semantics of object types. His system constitutes an extension of the theory of conceptual spaces. It primarily

addresses the limitations of classical (unrestricted extensional) modal logics by differentiating types that represent ascribed properties from those that carry a principle of identity (also known as sortal types). The system is exemplified, among others, by means of standard examples of "identity loss" (such as the alleged non-identity of a statue with the lump of clay constituting it), a notion he rejects. It is moreover shown how to circumvent some of the limitations that arise in representing modal (temporal) information with languages such as the Web Ontology Language (OWL) used in computer science.

Antonio Chella presents a cognitive architecture for a musical agent based on the architecture developed in Chella et al. (2001a) for computer vision. The underlying conceptual space is constructed from the two fundamental dimensions *pitch* and *time*. The timbre of a complex tone can then be represented as a vector of the harmonic frequencies about a fundamental tone. This representation may be used to generate a higher-level conceptual space in which the distance between tones based on their consonance can be determined, and musical intervals be represented. The new conceptual space forms the semantic base for a linguistic level of musical terminology. Chella also points out the multi-faceted analogies between vision and music perception.

1.2.3 Philosophical Perspectives

The theory of conceptual spaces also generates a number of philosophical issues. A fundamental question for epistemology, for instance, is how knowledge is best represented. The last part of the volume addresses some of these issues.

Lieven Decock and *Igor Douven* show conceptual spaces to be a particularly useful tool for philosophers. They discuss recent applications of the model to classical problems in metaphysics and the philosophy of language pertaining to identity such as the vagueness of terms, graded membership, and paradoxes of identity. They moreover summarize an analysis of knowledge that models the "usual suspects" (truth, belief, and justification) as a 3-dimensional space—*knowledge* being represented as one of its regions—to which additional epistemologically relevant dimensions can be added. As they show, the similarity between agents' doxastic states and the knowledge-region can thus be exploited to account for the observation that our willingness to attribute knowledge to people is stake-sensitive. Generally, as they stress, the "investigation of many concepts must involve more than an analysis of their component parts, and must involve some construction work as well: out of the component parts, a model of the concept must be built."

Addressing its testability, *Joel Parthemore* presents his unified conceptual spaces theory (UCST) and outlines various indirect ways of pitching it against empirical data by using mind-mapping software. This software is distinct from other available mind-mapping programs because it is directly based on a theory of concepts, rather than being loosely based on an underspecified theory of cognition. It also comes with a visual interface based on Voronoi tessellation. UCST offers a 'just so' story

of how all concepts describable within this framework can be derived from three proto-conceptual entities, and thus oriented within a unified "space of spaces" along one of three dimensions that are claimed to be integral to all concepts, namely the axes of *generalization*, *alternatives*, and *abstraction*. Within this unified space, the conceptual agent is assumed to build, modify, and navigate the conceptual frameworks that serve to structure and restructure an understanding of the world around her.

Mauri Kaipanen and *Antti Hautamäki* pursue an epistemologically grounded perspectivist approach to conceptual spaces. They present a model which comprises a multi-dimensional ontospace, a full grasp of which is limited by human cognitive capabilities, as well as a lower dimensional representational space that, next to supporting particular conceptualizations of the ontospace, allows for various alternatives thereof. Assuming that an ontospace is cognitively accessible only through the epistemic "work" of exploring alternative perspectives, they suggest that an understanding of a particular domain emerges only through having viewed it from multiple perspectives, thus abstracting further than any one given perspective. Such perspectives are said to vary individually as a function of interest, situational contexts, as well as various temporal factors; but they also remain communicable, and thus allow for interpersonally shared conceptualizations.

Frank Zenker and *Peter Gärdenfors* have in earlier work (Gärdenfors and Zenker 2013; Zenker and Gärdenfors 2013) shown how conceptual spaces apply to model changes of scientific frameworks (e.g., in physics), when these are treated as spatial structures, rather than as linguistic entities. In their chapter, this application is contrasted with Michael Friedman's (2001) neo-Kantian account, which particularly seeks to render the transition from Newtonian mechanics to relativity theory as a communicatively rational conceptual development. To compare different paradigms, Friedman introduces philosophical meta-paradigms as necessary elements. In contrast, Zenker and Gärdenfors argue that when theory frameworks are modeled as conceptual spaces and theories as constraints on them, then the communicative challenges said to go along with a paradigm shift become smaller. Thus, the cross-paradigmatic comparison of such frameworks that is necessary for rational scientific communication may instead proceed via the frameworks' geometric or topological properties. This, they argue, lies closer to what scientists in fact use in their thinking and communication.

Acknowledgements Peter Gärdenfors gratefully acknowledges support from the Swedish Research Council for the Linnaeus environment *Thinking in Time: Cognition, Communication and Learning*. Frank Zenker acknowledges funding from the Swedish Research Council.

References

Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In E. Kittay & A. Lehrer (Eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization* (pp. 21–74). Hillsdale: Lawrence Erlbaum Associates.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

Chella, A., Frixione, M., & Gaglio, S. (2001a). Conceptual spaces for computer vision representations. *Artificial Intelligence Review, 16*, 137–152.

Chella, A., Gaglio, S., & Pirrone, R. (2001b). Conceptual representations of actions for autonomous robots. *Robotics and Autonomous Systems, 34*, 251–263.

Friedman, M. (2001). *Dynamics of reason*. Stanford: CSLI Publications.

Gärdenfors, P. (1990). Induction, conceptual spaces and AI. *Philosophy of Science, 57*, 78–95.

Gärdenfors, P. (1992). A geometric model of concept formation. In S. Ohsuga et al. (Eds.), *Information modelling and knowledge bases III* (pp. 1–16). Amsterdam: IOS Press.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press.

Gärdenfors, P. (2007). Representing actions and functional properties in conceptual spaces. In T. Ziemke, J. Zlatev, & R. M. Frank (Eds.), *Body, language and mind, volume 1: Embodiment* (pp. 167–195). Berlin: Mouton de Gruyter.

Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. Cambridge, MA: MIT Press.

Gärdenfors, P., & Warglien, M. (2012). Using conceptual spaces to model actions and events. *Journal of Semantics, 29*, 487–519.

Gärdenfors, P., & Zenker, F. (2013). Theory change as dimensional change: Conceptual spaces applied to the dynamics of empirical theories. *Synthese, 190*, 1039–1058.

Jäger, G. (2010). Natural color categories are convex sets. *Amsterdam Colloquium 2009, LNAI 6042*, pp. 11–20. Springer: Berlin.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.

Langacker, R. W. (1987). *Foundations of cognitive grammar, vol. I*. Stanford: Stanford University Press.

Mervis, C., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*, 89–115.

Okabe, A., Boots, B., & Sugihara, K. (1992). *Spatial tessellations: Concepts and applications of Voronoi diagrams*. New York: Wiley.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*, 192–233.

Rosch, E. (1978). Prototype classification and logical classification: The two systems. In E. Scholnik (Ed.), *New trends in cognitive representation: Challenges to Piaget's theory* (pp. 73–86). Hillsdale: Lawrence Erlbaum Associates.

Schiffman, H. R. (1982). *Sensation and perception* (2nd ed.). New York: Wiley.

Steels, L., Kaplan, F., McIntyre, A., & Van Looveren, J. (2002). Crucial factors in the origins of word-meaning. In A. Wray (Ed.), *The transition to language* (pp. 252–271). Oxford: Oxford University Press.

Warglien, M., Gärdenfors, P., & Westera, M. (2012). Event structure, conceptual spaces and the semantics of verbs. *Theoretical Linguistics, 38*, 159–193.

Zenker, F. (2014). From features via frames to spaces. Modeling scientific conceptual change without incommensurability or aprioricity. In T. Gamerschlag, D. Gerland, R. Osswald, & W. Petersen (Eds.), *Concept types and frames. Applications in language and philosophy* (pp. 69–89). Dordrecht: Springer.

Zenker, F., & Gärdenfors, P. (2013). Modeling diachronic changes in structuralism and conceptual spaces. *Erkenntnis, 79*(8), 1547–1561.

Part II
Semantic Spaces

Chapter 2
From Conceptual Spaces to Predicates

Jean-Louis Dessalles

Abstract Why is a red face not really red? How do we decide that this book is a textbook or not? Conceptual spaces provide the medium on which these computations are performed, but an additional operation is needed: Contrast. By contrasting a reddish face with a prototypical face, one gets a prototypical 'red'. By contrasting this book with a prototypical textbook, the lack of exercises may pop out. Dynamic contrasting is an essential operation for converting perceptions into predicates. The existence of dynamic contrasting may contribute to explaining why lexical meanings correspond to convex regions of conceptual spaces. But it also explains why predication is most of the time opportunistic, depending on context. While off-line conceptual similarity is a holistic operation, the contrast operation provides a context-dependent distance that creates ephemeral predicative judgments ('this book is not a textbook', 'this author is a linguist') that are essential for interfacing conceptual spaces with natural language and with reasoning.

2.1 Introduction: Meaning *vs.* Predicates

The word 'concept' has been heavily used in semantics, despite its ambiguity (Machery 2009). Most authors in linguistics use it to designate 'lexical meanings'. They would speak of the concept of 'book' and may sometimes write it BOOK. Some authors continue by 'grounding' the concept in perception, considering that the concept refers to an actual object (a book) or, in the absence of any context, to a *prototype* of book. Advocates of the prototype approach would allow membership (being a book or not) to be gradual (Rosch 1978). Authors from the logical tradition or the philosophical tradition would rather consider the membership function BOOK(x) as having a binary value (true or false) and would call it a (logical) *predicate*. Advocates of this second approach will equate the concept with its 'extension', which corresponds to its so-called 'truth values' (the set of all x that make BOOK(x) true). Much misunderstanding results from the fact that the word

J.-L. Dessalles (✉)
Network and Computer Science Department, Telecom ParisTech, Paris, France
e-mail: dessalles@telecom-paristech.fr; http://www.dessalles.fr

F. Zenker, P. Gärdenfors (eds.), *Applications of Conceptual Spaces*,
Synthese Library 359, DOI 10.1007/978-3-319-15021-5_2

'concept' may be used by different authors to designate sometimes prototypes and sometimes predicates. The purpose of the present paper is to suggest that the two notions should be kept separate.

This position may be surprising, as predication, understood as concept membership, is often considered by both schools to rely on fundamental cognitive abilities, such as object recognition, that we share with other animals. Hence the proposal that animals do have concepts (Fodor 1975) and that there is full evolutionary continuity between non-human primates and humans in this aspect of the semantic competence (Tomasello 1999; Hurford 2003).

There are strong reasons, however, to consider that prototypes and predicates both exist as cognitive representations, but are different in nature. The suggestion will be that predicates are transient representations that are built 'on the fly' based on prototypes, or rather on the kind of conceptual representations hypothesized by Gärdenfors (2000, 2014). In the next section, I will oppose the two traditions mentioned above (prototypes *vs.* predicates) and highlight the fact that both, separately, are unsuccessful in solving the problem of lexical meanings. Then, I consider how conceptual spaces (Gärdenfors 2000) deal with the challenge of interfacing with reasoning. Building on that model, I will define the *contrast operation*, to show how conceptual spaces may support logical reasoning. As will be suggested, predicates are formed 'on the fly' and are ephemeral constructs. To illustrate this point, I will consider temporal aspect and its role in predication. Lastly, I will consider the predicative ability from a functional and evolutionary perspective.

2.2 Two Incompatible Definitions of Meaning

Meanings are attached to language: words, phrases and sentences can be meaningful. But meanings play other roles. They refer to perceived objects, scenes or events.

(1) The red book on the chair to the right of my desk

In (1), the phrase refers to a specific object that may be, for instance, wanted by the speaker. To be fully understood, the phrase must be processed in the current environment, in interaction with perception. One wouldn't say that the interpreter got the meaning of the phrase if she is unable to provide a spatial description of the scene, for instance by drawing a sketch of it. Meaning is not only in words; it is 'rooted', or 'grounded', in perception. The difficulty of assigning a meaning to (2) does not lie in the meaning of words, but in the difficulty of forming an image of the object in the absence of any particular context.

(2) The garden of the door

Meanings are not only related to perception, but also to reasoning. Language is not used just to refer to things in the environment, as in (1). It is used to convey propositional attitudes. Example (1) is incomplete in this respect and may generate an answer like: "So what?". Attitudes express surprise, (dis)belief, and positive

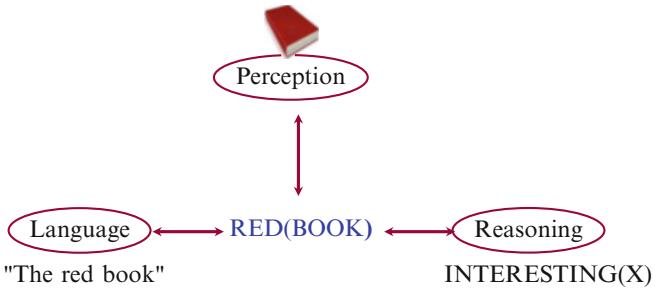

Fig. 2.1 The three interfaces of meaning

or negative emotions or desires (Bratman et al. 1988; Dessalles 2007). One may continue example (1) by saying "It's no longer there!", or "It's the best I've read recently". In the Fregean tradition, only *predicates* (with no free variables) can support attitudes. A basic test to recognize predicates is negation. A perceptual meaning of 'book' cannot be negated. No perception can illustrate what a *not-book* would look like. However, negation can be systematically used on predicates: "It isn't a book", "it's not red", "I haven't read it".

When defining the nature of meanings, one has to ensure that they correctly interface with these three domains: language, perception, reasoning (Fig. 2.1). Unfortunately, the corresponding requirements seem irreconcilable (Ghadakpour 2003). Perceptual representations such as prototypes *à la* Rosch (1978) or regions of 'conceptual spaces' *à la* Gärdenfors (2000) are of course perfectly fit for matching perceptions. They also interface well with words. Using figure-ground distinctions (Gärdenfors 2014), they can be used to provide different meanings to expressions like "the neighbour's cousin" and "the cousin's neighbour". However, they do not support negation and logical reasoning: there is no prototype for 'not-book' or 'not-red book'. Attempts to define negation as a comparison procedure would miss the point. The problem is to define a context-sensitive distance and a context-sensitive threshold to decide whether this object is, or is not, a book. No absolute distance from prototype can support judgments like "This is not a book, it is too thin", "This is not a book, it lacks bibliographic references" and "This is not a book, it's a collection of essays".

Another famous attempt to define meanings consists in considering that they are built on a 'mental lexicon'. According to this tradition, meaning construction boils down to a mere translation from actual language to a 'language of Thought' (LOT) (Fodor 1975). This approach has potential internal inconsistencies, though, depending on how the mental lexicon is constructed (Fodor 1981, 1998; Fodor and Lepore 1992; Ghadakpour 2003). One may choose to *define* lexical meanings in terms of other meanings. The meaning of 'kill' will be equated with something like KILL(x, y) ≡ CAUSE(x, DEATH(y)). Various formalisms have been proposed to encode such definitions, including flat logical expressions, description logics, λ-calculus, lexical conceptual structures, recursive feature structures and frames.

The very idea of definition is, however, problematic (Fodor et al. 1980; Fodor 1994). First, correct and complete definitions are nearly impossible to find (except in the limited domain of mathematics). For instance, one wouldn't say that a judge is killing the defendant by sentencing her to death, despite definitely causing her death by doing so. Second, definitions do not avoid the problem of having primitive, undefined meanings, such as CAUSE in certain accounts (Jackendoff 1983). Third, definitions offer no grounding in perception (Harnad 1990). They provide no means to distinguish geese from ducks (Jackendoff 1983) or pebble from stone or rock. Lastly, by making no distinction between the inner structure of lexical meanings and the structure of combined meanings attached to phrases, the definitional approach explains meaning construction by mere structural matching or unification; but meanings can only grow in this process, leading to implausibly large structures when discourse is processed. This problem has been called the *monotonous compositionality paradox* (Ghadakpour 2003).

Different studies in cognitive linguistics proposed dynamic mechanisms as an alternative to static definitions. For instance, force dynamics (Talmy 1988) intends to capture aspects of causation and modality. Note, however, that most of these approaches still postulate the existence of static structures attached to words, such as image schemas (Langacker 1987). For this reason, they are not immune to the critique addressed to the definitional version of LOT. The same is true of procedural approaches to meaning, such as frames (Johnson-Laird 1977), in which lexical meanings depend on procedures that must be executed on the fly. Despite their dynamic aspect at the time of execution, the procedures themselves constitute static structures attached to the lexicon. This makes them vulnerable to all the problems of the definitional version of LOT.

An alternative version of the LOT hypothesis considers that lexical meanings are atomic, *i.e.* have no internal structure, but are linked to each other. Different formalisms have been proposed to represent links between meanings, including logical knowledge bases, semantic networks, theories and conceptual graphs. This *relational* view of meaning is however exposed to the problem of holism and to the frame problem (Fodor and Lepore 1992): There is no way to circumscribe the effects of elementary changes in knowledge.

Both approaches to LOT, the definitional one and the relational one, are equally exposed to the grounding problem. Their representations are just floating symbols, defined using other symbols or being linked to them, without ever being connected to perceptions (Harnad 1990). The very idea of a LOT, based either on definitions or on relations, amounts to a clumsy duplication of perception. How can we decide which perceptions deserve being represented in LOT? Since perceptions are continuous and graded, whereas a LOT, like any language, is necessarily discrete, most perceptual nuances are supposed to remain outside of our conceptual power. The LOT hypothesis therefore does not explain why human beings can conceptualize any perceptual distinction. For instance, one may compare two colours and state that one is lighter than the other, without being able to define the two colours or even being able to name them.

Considering certain problems of the traditional approaches to LOT, Fodor proposed a non-definitional and non-relational approach, in which all concepts preexist, before some of them happen to be used (Fodor 1998; Chomsky 2000). This innatist model of conceptual knowledge is still vulnerable to the grounding problem. Moreover, it is forced to deny the very possibility of acquiring new meanings that were not pre-existent. Note that the definitional and the relational version of LOT are quite uncomfortable with the acquisition issue, as the child has to guess non-trivial (external or internal) structures when exposed to a new word.

Neither the prototype approach nor the LOT hypothesis seem to support the three interfaces that any theory of meaning must explain: with language, with perception and with reasoning. Prototypes do not support negation, and LOT meanings are not grounded. The only way out of the conundrum, as we suggest following Ghadakpour (2003), is to consider that there are no such things as complex permanent symbolic structures attached to the lexicon. The structural part of meanings that supports their logical role will be presented as an ephemeral construct. In the model presented here, predicates are only transitory and do not exist as permanent structures.

2.3 Conceptual Spaces and Categorization

Our problem, in a nutshell, is to define grounded predicates. Let's start by accepting Harnad's (1990) point that by combining purely symbolic structures, one will never get grounded representations. One may know that a book is a physical object made of paper that can be read, but if one has no perceptual grasp of what 'physical', 'paper' and 'read' mean, one will never be able to recognize actual books when one perceives them. Moreover, one will be unable to understand metaphors involving perceptual distance such as "It's not a letter you are writing, it's a book!" or "This cupboard opens like a book".

Our best chance is therefore to start from grounded representations and see how we can allow them to support negation, propositional attitudes and logical reasoning as predicates do. We may start from conceptual spaces (Gärdenfors 2000, 2014) because, as we will see, their geometrical nature makes them suitable to our purpose.

Gärdenfors (2000, 2014) offers an original account of the nature of meaning. He insists that meanings are geometrical entities. They belong to metric spaces that they share with perceptions. As a consequence, two meanings in the same 'conceptual space' may be more or less similar. More importantly, lexical meanings refer to convex regions in one of these 'conceptual spaces'. This means that if two objects are called 'book', all objects that fall in-between in the conceptual space will be identified by the word 'book' as well. Gärdenfors set himself the goal of identifying the various spaces in which lexical meanings are located, depending on the semantic nature of meanings (such as events, actions or qualities) or on their syntactic role (such as nouns, adjectives, verbs). He observes for instance that nouns correspond to regions in multi-dimensional spaces, whereas adjectives refer to low-dimensional or even one-dimensional domains.

Conceptual spaces, as they are described by Gärdenfors, support categorization, but not predication. Categorization means that when an object is perceived, it can be assigned to a known category. Since conceptual spaces are metric spaces, such categorization is straightforward: Just associate the object to the closest category represented by the exemplars and prototypes that have been memorized. This closest-neighbour device, when strictly applied, divides the semantic space in juxtaposed polygonal cells and produces a pattern called a Voronoi partition (Gärdenfors 2000: 97). Categorization, performed this way, differs however from predication.

First, categorization is a very basic operation that is far from representing the type of membership judgments performed by human beings. Even bacteria can be said to categorize, when they accelerate in acidic regions and slow down in neutral environments, or when they synthesize β -galactosidase only in the presence of lactose. Are we ready to say that bacteria categorize regions based on their acidity? That they possess the concept of acidity? Or the concept of lactose?

On the other hand, membership judgments, when performed by human beings, have two non-trivial properties: they are context-dependent and they can be justified. One may say for instance, talking about an electronic book: "It is a book, because it has been properly published". Though a prototypical book is still nowadays a physical object, a scrolling text on screen can be called 'book' without hesitation in an appropriate context. The membership judgment in this example contains a justification which is also context-dependent: "because it has been properly published". Negative membership judgments like "This is not a book, it is too thin", "This is not a book, it lacks bibliographic references" and "This is not a book, it's a collection of essays", contain context-dependent justifications as well. These justified positive and negative membership judgments are exactly what predicates do.

Distance-based categorization does not produce any negative membership judgment. Everything is a book, more or less. The only way to refuse bookness to an object would consist in finding a closer resemblance to another prototype, *e.g.* a notebook. But a thin book may look like a notebook and still be a book. Distance-based categorization cannot produce context-dependent negation. Moreover, convex regions of conceptual spaces do not support negation. Attempts to define 'not-book' and 'not-red' would produce non-convex regions. For instance, if 'not-red' is understood as 'any colour but red', it will correspond to the whole colour space with a convex hole in it and will not itself be convex. Moreover, standard distance-based categorization cannot give rise to any justification, other than "It looks like X". This mere resemblance statement given as justification cannot be context-dependent. As we can see, there is quite a gap between standard categorization and predication.

Conceptual spaces can obviously produce basic categorization, based on distance to nearest neighbour. However, Gärdenfors does not provide any mechanism through which conceptual spaces would produce predicative judgments. The next section shows how such a mechanism can be defined.

2.4 The Contrast Operation

The main thesis of the present paper is that membership judgments, which are sometimes considered to be the main role that 'concepts' have to fulfill, cannot be deduced from distance to prototypes. Available distances, such as those used in prototype theory, in neural networks or in conceptual spaces, are *holistic*. This means that they are computed on many dimensions: all available dimensions in prototype theory, dimensions that are statistically relevant in neural networks, or the various dimensions of the conceptual space to which the prototype belongs. An object that differs from the prototype in only one respect is unlikely to be rejected using a holistic distance. A negative membership decision like "This object is not a book, it is too thin" will not be generated if the object has all the typical characteristics of a book but thickness. Discrepancy on only one dimension does not make a difference when there are many dimensions (colour, shape, matter, pages, printed content, title, publisher, references, ...) on which the object matches the prototype.

Following Ghadakpour (2003), we claim that membership judgments are generated by using a *contrast operation*. The object O is contrasted with the *closest* prototype P. We may write the output of this operation $C = O - P$. The use of the minus sign is not fortuitous: in the spirit of conceptual spaces, the contrast operation can be implemented using the difference between vectors. Note that C is not a distance, but a vector (*i.e.* a concept). Figure 2.2 illustrates how contrast works. The representation of a given book O is contrasted with the 'textbook' prototype. The resulting vector, shown as a dashed arrow, provides a *contrastive dimension*.

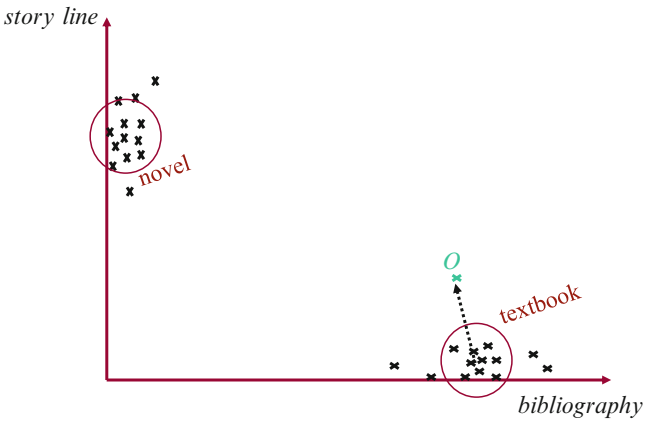

Fig. 2.2 Schematic representation of the contrast operation. *Circles* represent two prototypes (novel and textbook) along two dimensions: quantity of bibliographic references and coherence of the story line. *Dots* represent a few books that are close to the two prototypes. The *dashed arrow* represents the contrastive dimension between book O and the 'textbook' prototype

Though Fig. 2.2 shows only two dimensions, the contrast operation is performed in a multi-dimensional space including all dimensions on which objects (here, books) are perceived.

The contrastive dimension serves as basis for several further operations. One of them is *modification*. For instance, as far as the contrastive dimension matches the 'story line' dimension, O may be characterized as 'a textbook with a story line', or as a 'narrative textbook'. Since O and the associated prototype are close along many dimensions, C can often be approximated by a low dimensionality vector. For this reason, according to Gärdenfors, C is likely to fall in the region of an adjective. If one compares a 20-page book to a prototypical book, the difference will be close to the prototype of 'thin' or 'thinness'.

Modification through adjectives ('narrative', 'thin') or through specifying phrases ('with a story line') is involved in the operation of predication, and translates directly into logical predicates ($thin(O)$, ...). These predicates, which result from contrastive judgments, are not easily produced by systems that rely exclusively on global distances, such as prototype-based models or conceptual spaces in their basic form, without introducing external notions such as 'salience'. Modifying predicates, however, come naturally with the contrast operation. In a judgment like "This girl has a red face", the adjective "red" needs not be metaphorical nor a distorted version of the redness prototype. It merely results from the contrast between the girl's face and a prototypical face, as the difference may happen to match the prototype of 'red'. The modifier ('narrative', 'with a story line', 'red') is not a distorted representation, but an accurate representation of the difference $O-P$ (Fig. 2.2).

The contrast operation, as proposed here, can be seen as a cognitive device that supports 'lexical contrast' (Murphy 2003: 170). Moreover, it avoids any notion of 'salience' to explain context dependency. Adjectives like 'thin' or 'big' may be applied to objects of various sizes: "a big flea", "a big galaxy". The meaning of these phrases cannot be determined by two separate computations, as in a writing like $\mathrm{BIG}(x)$ & $\mathrm{FLEA}(x)$ (Kamp and Partee 1995). The meaning of 'big' has to be understood in the flea context or in the galaxy context. The contrast operation provides this context dependency by determining both that the adjective 'big' is relevant and on which scale it is interpreted. The perceived flea (resp. galaxy) contrasts with the prototype of flea (resp. of galaxy) by its size; the scale (millimetres *vs.* thousands of light-years) comes from the standard deviation of the prototype. This works if the prototype of 'big' belongs to the conceptual space of homothetic transformations.

The contrastive dimension plays also an essential role in the most basic form of predication: membership judgments. In the example of Fig. 2.2, should we opt for $textbook(O)$ or for its negation? The contrast operation by itself does not produce membership judgments. A binary decision is required in addition. The contrast operation provides the space on which this membership decision will be made. "This book is thin, but it is still a book", or "This book is too thin to be a proper book". This kind of binary decision is a *topological* decision. It first creates a frontier F on

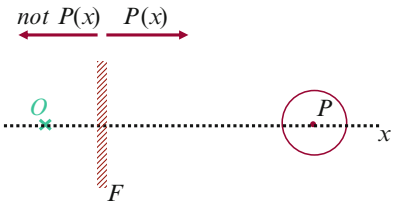

Fig. 2.3 Membership decision based on contrast

the contrasting dimension, and then decides whether the object and the prototype are on the same side of the frontier or on opposite sides (Fig. 2.3).

The combination of contrast and binary separation has been invoked to account for antonym pairs (Paradis and Willners 2011). By taking 'contrast' literally as a difference, however, we explain context sensitivity in oppositions ('chair *vs.* stool', 'chair *vs.* sofa') without appealing to notions such as 'contextually relevant properties'. Relevance is an output, rather than an input, of the contrast operation (Dessalles 2008).

Interestingly, the combination (contrast + binary decision) provides additional support to Gärdenfors claim about the convexity, or at least the star-like nature of concepts, considered as regions in a conceptual space (Gärdenfors 2000: 69). If an object is found to fall inside the frontier, any other object that is closer to the prototype along the contrasting dimension will be located inside as well. The same reasoning can be made about any contrasting dimension. As a consequence, we expect membership judgments to generate star-shaped regions around the prototype. The observation that meaning should self-organize to produce a Voronoi partition (Gärdenfors 2000: 91) may follow from this star-like property, but only indirectly. Gärdenfors' observation refers to average lexical meanings emerging from language use within a linguistic community. The contrast operation, however, produces only situated judgments, for one speaker at a given moment. It is performed dynamically, on one dimension at a time. For convexity to emerge in conceptual spaces, judgments must be consistent across speakers, at least on average.

The present model makes a clear distinction between prototypes and predicates. The former are long-lasting representations. The latter are ephemeral representations. They are produced 'on the fly', possibly with insincerity as their purpose is often to make a point, as in "This is not a book, it's a collection of essays". Predicates support negation and attitudes, and are used in reasoning. They differ from prototypes, but are built on them. The fact that we can perform membership judgments on any object–prototype pair gives the illusion that predicates are permanent representations. They are not. They are created when needed, and they vary according to contexts. Similarly, our ability to impose binary decisions on any contrasting dimensions, graded or not, gives the illusion that 'concepts' can be defined. But these 'definitions', which serve as basis for most LOT approaches to semantics, are just-so constructs that are made up on the fly. They are not stored permanently (except in the rare cases in which their linguistic form is memorized, as when a student learns definitions by heart) (Ghadakpour 2003). For instance, anyone

knows the difference between the meanings of 'walk' and 'run', but few people are able to provide a definition for these terms. In race walking competitions, 'walking' obeys two constraints: both feet cannot lose contact with the ground simultaneously, and the supporting leg must straighten from the point of contact with the ground and remain straightened until the body passes directly over it. Such fixed definitions are no more than *ad hoc* conventional constructs. They have no cognitive reality, except for the very few that have been learned by heart (Fodor et al. 1980). The only permanent conceptual features are perceptual, and are well modelled by prototypes or exemplars located in conceptual spaces, as described by Gärdenfors.

The predicative ability shows up in various parts of our linguistic competence. I mentioned membership judgments, negation and attitudes. The contrast operation is also involved in comparatives such as "warmer" and "taller". No wonder, since contrast is likely to produce low-dimensionality output, that comparatives are generally expressed through adjectives, as observed by Gärdenfors (2014). A comparative adjective like 'first' is ambiguous: "The first 4G-network" may mean the first by its geographic coverage or the first to be deployed. This ambiguity is best explained by the fact that the contrast operation can produce different dimensions, depending on the context.

This last example highlights the fact that contrast may operate on any dimension, including time. Contrasting operations are even essential for generating aspectual distinctions. This is what we explore in the next section.

2.5 Predication and Temporal Aspect

From a cognitive perspective, the most basic aspectual property is the notion of *perfectivity*. This notion corresponds to the fact that a situation may be perceived and expressed either as bounded (perfective) or as unbounded (imperfective). Since perfectivity may change during semantic processing, as we will see, I prefer to use the notion of *viewpoint* (Ghadakpour 2003; Munch and Dessalles 2014; Munch 2013). A situation can be seen from the outside, in which case it is seen as a (bounded) *figure*, or from the inside, in which case it is seen as an (unbounded) *ground*. The notions of 'figure' and 'ground', borrowed from Gestalt psychology, seem appropriate here, as they avoid any idea of border, contour or limit. A 'figure' is a whole (with neither interior nor frontier) and a 'ground' is regarded as a limitless area. The figure/ground distinction associated to perfectivity is illustrated in French by the difference between present perfect ("Elle a mangé") and the imperfective ("Elle mangeait"). Incompatible viewpoints provoke semantic errors.

(3) She wants to eat the entire cake in 1 min
(4) # She wants to eat the entire cake for 1 min

Example (3) is correct. Example (4), however, is hardly acceptable. "To eat the (entire) cake" is a figure. It is compatible with the figure introduced by 'in' ("in one minute") but it does not match the ground introduced by 'for' ("for one minute").

In examples (5) and (6), we can observe the converse conflict, as "to snore" is perceived as a ground ('snoring' refers to a homogeneous situation).

(5) # She is expected to snore in 1 h (tomorrow)
(6) She is expected to snore for 1 h (tomorrow)

Note that (5) would be acceptable with a meaning like (7), which is sometimes called 'inchoative'.

(7) She is expected to snore after 1 h (tomorrow)

The effect of 'in' would be the same as in "He would confess his crime in one hour (if I could question him)". Similarly, (8) is perfectly acceptable.

(8) She is expected to snore in 1 h (from now)

In (7) and (8), the mention of duration concerns not the situation itself (snoring), but the period that precedes it. The simplest explanation of this apparent inchoativity consists in the fact that 'snore' denotes a situation in (5) and (6), but is predicated in (7) and (8). In these latter sentences, the 'snoring' situation is contrasted with a 'non-snoring' situation. Both 'snoring' and 'non-snoring' become figures that are topologically separated on a snoring gradient space. This change of conceptual space means also that 'snoring' looses its temporal dimension (Munch and Dessalles 2014; Munch 2013). A meaning cannot be contrasted on two different conceptual spaces at the same time. Being now atemporal, the 'snoring' situation cannot be matched with 'in one hour'. The English language allows the atemporal event to shift towards the end of the period introduced by 'in'.

The loss of temporality due to predication can also be seen in the two following examples.

(9) She intends to drink champagne next year (to celebrate her election)
(10) She intends to drink champagne next year for 1 h

In (9), 'to drink champagne' has the meaning of 'to celebrate' or even, metonymically, of 'to be elected'. This is a predicative use, where the contrast is 'drink' *vs.* 'not drink'. In (10), temporality is imposed by the mention "for one hour". The predicative use of 'drink champagne' is now excluded. 'drink' recovers its concrete meaning of activity and can no longer designate the act of celebrating.

Predication plays a prominent role in the determination of aspect. The phrase 'to eat cookies' normally corresponds to a ground, as it refers to a repetitive activity. But in (11), it plays the role of a binary action. This binary nature results from predication: 'to eat cookies' is contrasted with its negation (Fig. 2.4). It has therefore no temporal dimension. For instance, she was supposed to fast, but she broke her vow. Under this binary interpretation, 'to eat cookies' is now a figure. This allows the perfective aspect indicated by the present perfect.

(11) I know that she has eaten cookies

As we can see, the contrast operation on which predication relies provokes a change of conceptual space. The resulting situation must be considered as having no duration whatsoever (Fig. 2.4).

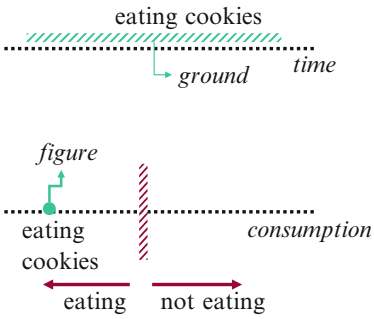

Fig. 2.4 Loss of temporality after contrast. When 'eating cookies' is contrasted with its negation, 'not eating cookies' (*lower diagram*), it is no longer a temporal ground (*upper diagram*), but a figure

The existence of this mechanism shows that most situations are not by themselves figures or grounds. For instance, activities and states are classically distinguished from achievements and accomplishments (Vendler 1967). Both accomplishments (to eat the entire cake) and achievements (to sneeze) are systematically regarded as figures (*i.e.* they are perfective), whereas states and activities are systematically supposed to be grounds (they are imperfective). The process of predication through contrast shows, however, that the distinction is not 'hard-wired'. Any state or activity, once predicated, becomes an atemporal figure. It then behaves as if it were instantaneous (Fig. 2.4). In (12), 'be happy' must be turned into a figure to match the figure introduced by 'in'. Predication does the job, but moves the 'happiness' state to another space, a happiness gradient, where it looses its temporality. As a consequence, (12) cannot mean that her happiness lasted for 1 month. Only the inchoative meaning is available (to become happy after 1 month).

(12) She wanted to be happy in 1 month. (= after 1 month)

We understand why predication provokes a conceptual space change in these examples. The contrast operation is performed on a conceptual space, which is not the temporal one (Fig. 2.4). One should not conclude that the output of predication is systematically atemporal. The binary interpretation of (11) is impossible in (13), where 'to eat cookies' must keep its temporal dimension.

(13) I know that she has been eating cookies for 1 h

But "eating cookies for one hour" is a ground. It does not match with the figure imposed by the present perfect. Predication can again do the job of turning it into a figure. How? By performing contrast, not on a conceptual dimension, but on duration. The binary distinction that supports predication is now between 'less that an hour' and 'more than an hour'. As a consequence, the propositional attitude in (13) must be toward the temporal extension ('more than an hour'), *e.g.* if eating cookies for that long is a feat. The same phenomenon applies in examples (3), (6) and (10). Note that in (3), the predication of "to eat the entire cake in one minute" is achieved through the converse contrast, this time 'less than one minute' against 'more than one minute'.

2.6 The Origin of Predicates

The central thesis of this chapter is that there is a fundamental difference between lexical meanings and predicates. Lexical meanings refer to exemplars and prototypes, which belong to conceptual spaces. Predicates, on the other hand, are produced on the fly through a contrast operation. The necessary use of contrast when performing membership judgments contributes to explaining why the 'extension' of lexical meanings in conceptual spaces is found to correspond to convex regions, as observed by Gärdenfors (2000).

This position entails that predicates are exclusively dynamic representations that have no long-term, context-independent existence. This solves the three-interface problem (Fig. 2.1). Predicates are no duplicates of perception, since they are built on perception when needed. And the fact that predicates are not permanent structures avoids all the paradoxes associated with the LOT hypothesis (Ghadakpour 2003): absence of correct and complete definitions, holism, grounding problem, monotony of compositionality.

The difference introduced here between meanings and predicates allows us to consider the possibility that we share the former with other animals, but not the latter (Dessalles 2007). Animals like chimpanzees are perfectly able to learn lexical meanings (Savage-Rumbaugh and Lewin 1994). There ability to form predicates is however doubtful. Animals can be trained to perform contrasts and to name them on specific dimensions. For instance, a grey parrot (Alex) was able to tell the difference between two collections of objects (Pepperberg 1999). Alex could say that the two collections differed by their shape or their colour. He could even say 'None' when he noticed no difference. This is exactly what our predicative ability allows us to do. Note, however, that Alex's performance results from hundreds of repetitive exercises and cannot be transferred to new conceptual spaces. It is likely that this performance, which can be reproduced with artificial neural networks, is obtained by statistical selection of relevant dimensions. The claim is not that animals cannot perform complex distinction between objects. They obviously can. But the distinction has no predicative status. The output of contrasts performed by human beings is a new conceptual representation that can be named. Every human being does this spontaneously on any perceptual dimension, without previous training. This ability, universal in our species, seems inexistent in other species.

If this hypothesis is correct, then we must ask what type of function the predicative ability has in our species that it has not in other species. I submitted elsewhere that predication emerged as a device to detect lies and errors (Dessalles 1998, 2007), and to do it publicly. The ability to contrast what others say with what we saw allows us to name the difference ("What she wrote is too *thin* to be a book"). In other words, predication would have emerged in the first place because it supports our ability to perform explicit negation. It is perfectly possible to process perception and even to communicate about it without the predicative ability. This is possibly what previous hominin species did if protolanguage, as defined by Bickerton (1990), ever existed. The combination of lexical meanings ("house + neighbour + fire") can

be done without any need of predication. However, negation, logical reasoning, comparisons and the expression of attitudes are achieved in our species by means of predicates.

We only considered here predicates with one variable. Though some authors claim that all predicates can be constructed based on one-place predicates (Hurford 2003), the way events and actions undergo predication is not yet clear from a cognitive perspective. The Gärdenfors (2014) model of events, which involves two entities (agent and patient) and two vectors (one representing the force exerted by the agent on the patient, the other to represent the resulting change), is an interesting step towards an explanation of how complex predicates are dynamically formed. Again, we must draw a clear line between lexical meanings (especially the meaning of verbs), which are prototypes, and predicates, which are built on the fly. The role of the contrast operation in such a process is still unclear.

By distinguishing lexical meanings from predicates, one avoids many of the traditional theoretical difficulties in which fundamental semantics is entangled. The acknowledgment of the dynamic nature of predicates, which are no more than transitory representations, opens the way to new approaches to some semantic phenomena. The case of aspectual properties has been evoked in the preceding section. We need to investigate other dynamic devices that lie at the interface between language, perception and reasoning (Fig. 2.1) and that allow us to process spatial relations, tense, determination, quantification or modality. The contrast operation, which operates on conceptual spaces, could play a significant role in these future theories.

Acknowledgments This research is based on past collaborations with Laleh Ghadakpour and with Damien Munch. I would like to thank Damien Munch for his fruitful comments. Part of this research is funded by the "Chaire Modélisation des Imaginaires, Innovation et Création" (http://imaginaires.telecom-paristech.fr).

References

Bickerton, D. (1990). *Language and species*. Chicago: University of Chicago Press.

Bratman, M. E., Israel, D. J., & Pollack, M. E. (1988). Plans and resource-bounded practical reasoning. *Computational Intelligence, 4*(4), 349–355.

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.

Dessalles, J.-L. (1998). Altruism, status, and the origin of relevance. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: Social and cognitive bases* (pp. 130–147). Cambridge: Cambridge University Press.

Dessalles, J.-L. (2007). *Why we talk – The evolutionary origins of language*. Oxford: Oxford University Press.

Dessalles, J.-L. (2008). *La pertinence et ses origines cognitives – Nouvelles théories*. Paris: Hermes-Science Publications.

Fodor, J. A. (1975). *The language of thought*. Oxford: Harvard University Press.

Fodor, J. A. (1981). *Representations: Philosophical essays on the foundations of cognitive science*. Cambridge, MA: MIT Press.

Fodor, J. A. (1994). Concepts: A potboiler. *Cognition, 50*, 95–113.
Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford: Clarendon.
Fodor, J. A., & Lepore, E. (1992). *Holism – A shoper's guide*. Cambridge, MA: Blackwell.
Fodor, J. A., Garrett, M. F., Walker, E. C. T., & Parkes, C. H. (1980). Against definitions. *Cognition, 8*, 263–367.
Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press.
Gärdenfors, P. (2014). *Geometry of meaning – Semantics based on conceptual spaces*. Cambridge, MA: MIT Press.
Ghadakpour, L. (2003). *Le système conceptuel, à l'interface entre le langage, le raisonnement et l'espace qualitatif: vers un modèle de représentations éphémères*. Thèse de doctorat, Ecole Polytechnique, Paris.
Harnad, S. (1990). The symbol grounding problem. *Physica D, 42*, 335–346.
Hurford, J. R. (2003). The neural basis of predicate-argument structure. *Behavioral and Brain Sciences, 26*(3), 261–283.
Jackendoff, R. (1983). *Semantics and cognition*. Cambridge: MIT Press.
Johnson-Laird, P. N. (1977). Procedural semantics. *Cognition, 5*, 189–214.
Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition, 57*(2), 129–191.
Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites*. Stanford: Stanford University Press.
Machery, E. (2009). *Doing without concepts*. Cambridge, MA: Oxford University Press.
Munch, D. (2013). *Un modèle dynamique et parcimonieux du traitement automatisé de l'aspect dans les langues naturelles*. PhD dissertation, *to appear* Telecom ParisTech 2013-ENST-0058.
Munch, D., & Dessalles, J.-L. (2014). Assessing parsimony in models of aspect. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the Cognitive Science Society* (pp. 2121–2126). Austin: Cognitive Science Society.
Murphy, M. L. (2003). *Semantic relations and the lexicon*. Cambridge: Cambridge University Press.
Paradis, C., & Willners, C. (2011). Antonymy: From convention to meaning-making. *Review of Cognitive Linguistics, 9*(2), 367–391.
Pepperberg, I. M. (1999). *The Alex studies – Cognitive and communicative abilities of grey parrots*. Cambridge, MA: Harvard University Press, ed. 2000.
Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale: Lawrence Erlbaum Associates.
Savage-Rumbaugh, E. S., & Lewin, R. (1994). *Kanzi: The ape at the brink of the human mind*. New York: Wiley.
Talmy, L. (1988). Force dynamics in language and thought. *Cognitive Science, 12*(1), 49–100.
Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.
Vendler, Z. (1967). *Linguistics in philosophy*. Ithaca: Cornell University Press.

Chapter 3
Conceptual Spaces at Work in Sensory Cognition: Domains, Dimensions and Distances

Carita Paradis

Abstract This chapter makes use of two data sources, terminological schemas for wine descriptions and actual wine reviews, for the investigation of how experiences of sensory perceptions of VISION, SMELL, TASTE and TOUCH are described. In spite of all the great challenges involved in describing perceptions, professional wine reviewers are expected to be able to give an understandable account of their experiences. The reviews are explored with focus on the different types of descriptors and the ways their meanings are construed. It gives an account of the use of both property expressions, such as *soft, sharp, sweet* and *dry* and object descriptors, such as *blueberry, apple* and *honey*. It pays particular attention to the apparent cross-sensory use of descriptors, such as *white aromas* and *soft smell*, arguing that the ontological cross-over of sensory modalities are to be considered as symptoms of 'synesthesia' in the wine-tasting practice and monosemy at the conceptual level. In contrast to the standard view of the meanings of words for sensory perceptions, the contention is that it is *not* the case that, for instance, *sharp* in *sharp smell* primarily evokes a notion of touch; rather the sensory experiences are strongly interrelated in cognition. When instantiated in, say SMELL, *soft* spans the closely related sense domains, and the lexical syncretism is taken to be grounded in the workings of human sensory cognition.

3.1 Introduction

In his book, *Remarks on Color* Wittgenstein (1977: 102) notes that "When we're asked What do 'red', 'blue', 'black', 'white', mean? we can, of course, immediately point to things which have these colors—but that's all we can do: our ability to explain their meaning goes no further". The same is of course true of the other sensory perceptions with one very important difference, namely that there is nothing to point to, suggesting other means of identification and description. But, what are they?

C. Paradis (✉)
Centre for Languages and Literature, Lund University, Lund, Sweden
e-mail: carita.paradis@englund.lu.se

F. Zenker, P. Gärdenfors (eds.), *Applications of Conceptual Spaces*,
Synthese Library 359, DOI 10.1007/978-3-319-15021-5_3

This chapter is concerned with how visual, olfactory, gustatory and tactile experiences translate into language and what conceptual structures are at work in this process. For this purpose, the study makes use of the same theoretical framework that has been used to describe and explain language use and meaning in 'non-technical' everyday language contexts, namely Lexical Meaning as Ontologies and Construals, LOC for short (Paradis 2005). This framework is couched in the broad framework of Cognitive Linguistics (e.g. Langacker 1987; Talmy 2000; Croft and Cruse 2004) and it shares crucial modeling aspects with Gärdenfors (2000, 2014) more interdisciplinary oriented cognitive semantic approach. The basic assumption of Cognitive Linguistics is that concepts develop out of bodily experiences in the cultural settings where speakers happen to be born and live. The way we perceive the world is the way we understand it and we express ourselves accordingly. This is what is referred to as embodiment in the Cognitive Linguistics literature. In actual fact, as it has been used there, embodiment is a holistic notion. Not much effort has been made to examine the various different sensory experiences. The role of VISION — more precisely, how humans view things and how this is revealed in how we express ourselves, has been given most attention.

Meaning in language is a viewing arrangement filtered through the eyes of the conceptualizer (Langacker 1999: Chapters 7 and 10, Beveridge and Pickering 2013). This makes questions about the resources that language offers, not only for VISION, but also for SMELL, TASTE and TOUCH very pertinent. A provocative proposal for a Sensitive Linguistics, in place of Cognitive Linguistics, has been expressed by anthropologist David Howes (2013: 298) in a postscript of a book entitled *Sensuous cognition*. Howes notes that it is about time that perceptions received the attention that they deserve. Up to now, not much effort has been made in Cognitive Linguistics to try to tease out the respective roles of different sensory perceptions. Howes is of the opinion that the overriding emphasis of the holistic view of the embodied mind has hampered research on the individual sensory perceptions, their interrelations and how they are talked about in different languages and different cultures. He concludes:

> Fortunately, the veil cast by the embodied mind theory is now being lifted, and we are beginning to see how the senses have minds of their own. To put this another way, the embodied paradigm is being *out-moded*, as more and more scholars [. . .] come to (and into) their senses and lay the foundations or a new science of *sensuous cognition* (in place of embodied cognition) or *Sensitive Linguistics* (in place of Cognitive Linguistics), which is attuned to the varieties of sensory expression and experience across cultures. (Howes 2013: 298)

It is precisely to this concern that the current chapter is devoted. It proceeds from sensations to language through cognition and investigates the lexical forms and their conceptual underpinnings. The contribution of this chapter is to shed new light on the relation between language, cognition and the sensorium on the basis of the following three questions.

- What kind of descriptors are there?
- What conceptual (ontological) structures are evoked in the descriptions of different sensory experiences?
- How does sensory cognition shape the language of perceptions?

Two types of data are used for the analysis: One type comes from terminological schemas for wine descriptions and the other from a database of wine reviews from the *Wine Advocate*, a wine magazine run by world famous Robert Parker.[1] The main motivation for using these two data sources is that while the former sheds light on how sensory perceptions *can* be translated into language within a system of analytic terms, the other data set is about how sensory perceptions *are* expressed through language on the basis the reviewers' tasting experience in order to be understood by the readers and hopefully also to evoke his or her sensorium. Wine reviews are useful as a source of information because they almost always cover four different senses (VISION, SMELL, TASTE and TOUCH, i.e. TEXTURE), and also because SMELL receives particularly detailed attention, which is interesting since there is an alleged paucity of vocabulary in language in this domain. The database is large, including almost 85,000 reviews, and therefore allows for principled computational techniques, instead of relying on more *ad hoc* data collection methods. It is worth noting in this context that these reviews have had an enormous impact in the wine world, not only among connoisseurs but among producers and retailers (Hommerberg 2011: 3–12; Hommerberg and Paradis 2014).

3.2 Transforming Sensations into Language

Like most experiences of the world, wine tasting experiences are highly complex interactions between sensory experiences and knowledge about entities in the world that give rise to sensorial responses. In the social and discursive practice, it is the task of wine critics to communicate their experiences, the success of which hinges on their ability to translate their sensations into language so that their readers can interpret their message, and, if possible, also that it gives rise to an aesthetic response upon reading about the wines. For this reason, it is natural to start at the practical end. The practical procedure is described by Gluck (2003) as follows.

> You pour out the wine. You regard its colour. You sniff around it. You agitate the glass to release the esters of the perfume and so better to appreciate the aromas, the nuances of the bouquet. You inhale those odoriferous pleasantries, or unpleasantries, through the chimney of the taste, the nostrils (the only access to the brain open to the air) and then you taste.

[1] I am grateful to Mr Robert Parker for providing the database which facilitated the work immensely (http://www.erobertparker.com/members/home.asp). I am also grateful to Mats Eeg-Olofsson who carried out the computational work and made the relevant searches. A description of the corpus can be found in Paradis and Eeg-Olofsson (2013). The database has also been the basis for the creation of an interactive visualization tool (Kerren et al. 2013).

> You swill the liquid around the mouth and breathe in air so that this liquid is aerated and experienced by up to ten thousand taste buds. The taste buds are arranged in sectors of differently oriented cohesion: one designed to recognize salinity, another alkalinity, another sweetness and so on. They connect with the brain which in turn provides the sensory data, memory based, to form the critic's view of what s/he is drinking. Some of the wine is permitted to contact the back of the throat, but only a small amount is permitted to proceed down the gullet, so that the finish of the wine can be studied. Then the wine is ejected and several seconds are left to elapse whilst all these sensations are studied and written up as the impression the wine has left is mulled over. (Gluck 2003: 109)

As described by Gluck (2003), the procedure includes five stages: First, the taster considers the visual impression of the wine, second the taster concentrates on the smell of the wine, its nose, and third, its taste and texture (touch) are evaluated. Stage four concerns the "internal" olfactory stage where the wine's aftertaste is assessed, and finally stage five deals with the finish, i.e. how the wine vaporizes. The wine tasting practice takes the visual experience as its point of departure and in this sense VISION is in a special position compared to the other senses. The visual properties of the wine can be observed without interference of other sensory input. Physiologically, vision is also known to be our most consistent source of 'objective' data about the world, whereas smell is noted to be an elusive phenomenon from a cognitive point of view, and to appeal strongly to our emotions (Classen et al. 1994: 2–3). Zucco (2007: 161) notes that communication among humans about olfactory perception is complicated by the fact that people are conscious of smells only when these are present, and it is not possible to retrieve olfactory stimuli from memory, unless a specific smell is there as a memory trigger. Needless to say, these characteristics apply to taste and feeling as well. This suggests similarities across the modalities, and as noted by Paradis and Eeg-Olofsson (2013: 38), it is not possible to "taste something without smelling something and we cannot taste something without feeling something and over and above everything is the sight of something".

Proceeding now from the actual sensations and practical procedure to language and rhetorical structure, we note that the wine review (1) is iconic with praxis in that it runs from the taster's inspection of the wine's visual appearance through smelling, tasting and feeling its texture, i.e., from VISION through SMELL, TASTE and TOUCH, as shown in (1).

(1) The 1996 Cabernet Sauvignon Madrona Vineyard is the most promising wine Abreu has yet produced. *The color is a murky opaque purple, suggesting extraordinary richness. The wine's forward, sweet berry-scented aroma includes hints of cassis, lead pencil, and licorice. Thick and rich, with the 1996 vintage's sweet tannin in evidence, this full-bodied, powerful yet gorgeously layered and pure* Cabernet Sauvignon will be more precocious and flattering at an earlier age than either the 1995 or 1994. It will have two decades of positive evolution.

The italicized middle part of the text describes the sensorial experiences (my italics). The tasting practice and the rhetorical structure go hand in hand. The color of this wine is described as *murky opaque purple.* It has a *sweet berry-scented aroma* including *hints of cassis, lead pencil, and licorice.* The taste and

texture are described as *thick and rich*, *with sweet tannin, full-bodied,* and *powerful yet gorgeously layered and pure*. While this is the part of the text which is in focus in this chapter, it also deserves to be pointed out that most wine reviews consist of three parts starting with production facts and ending with a concluding assessment and recommendation of prime drinking time (Caballero 2009; Paradis 2009; Hommerberg 2011; Caballero and Paradis 2013; Paradis and Hommerberg in press; Hommerberg and Paradis 2014). These parts may also include assessments, very often of a more holistic type such as *is the most promising wine Abreu has yet produced* in (1).

3.3 Lexical Meaning as Ontologies and Construals

The foundation of the approach to meaning in language that this study is based on is socio-sensory-cognitive. Meaning in language is deeply rooted in our experiences of the world around us and shaped by our perceptions and cognitive abilities. Language evokes and construes conceptual structures according to the required discursive and social intentions, actions and requirements (Paradis 2005, in press). This makes language modeling essentially the opposite of mathematics, which exclusively deals with the relations of concepts to each other without consideration of their relation to experience. The basic assumption of LOC is that concepts are firmly grounded in perception and our experiences of the world (e.g. Langacker 1987; Gibbs 1994; Talmy 2000; Barsalou 2008; Lacey et al. 2012; Gärdenfors 2014). Gärdenfors (2014: 15) notes that "[n]ot only can we talk about what we see, but we also see (and hear, etc.), in our inner worlds, what we talk about. Language and perception are communicating vessels: I regard this as one of the main foundations for semantics". This foundational assumption gains support in neurobiological works which shows that conceptual representations involve multiple levels of abstraction from sensory, motor and affective input, and that activation of these modalities is influenced by factors such as contextual demands, frequency and familiarity (e.g. Binder and Desai 2011).

Knowledge of language involves the coupling of a conceptual structure with a lexical form, e.g. WINE/*wine*. The concept WINE rests upon a complex web of concepts in different domains of knowledge. The relative salience of the various domains depends on when, how and why the word *wine* is used. In other words, knowledge of the meaning of a word involves the coupling of a form with a graded structure in conceptual space on the occasion of use in human communication. All language elements are triggers of conceptual portions from the total use potential, which has been built up over time from experience with language usage in different social and cultural settings (Paradis 2005; Tomasello 2003, 2008).

For instance, the meaning potential of *wine* involves conceptual structures in all kinds of different domains of knowledge, not only VISION, SMELL, TASTE and TOUCH, but its domain matrix also comprises knowledge structures such as VINTAGE, BARREL, VINEYARD, TERROIR, GRAPE, CELLARING, AGRICULTURE,

WINE SHOP, GLASS, WINE DISTRICT, OENOLOGY, ALCOHOL, VITICULTURE, PRICE, CONSUMER, PRODUCER, NUTRITION and so on and so forth. In the case of wine reviews, for instance, the relative salience of the various meaning structures differs in the above-mentioned parts of the texts. While vineyards and grapes are the focus of attention in the part concerned with the production of the wine, color, smell, taste and touch are important in the description of the cellaring and maturation in the recommendation. The framework of lexical meaning, LOC, states that meanings are not inherent in words as such but evoked by words. Meanings of words are always negotiated and get their definite readings in the specific contexts where they are used (Cruse 2002; Paradis 2005, 2008; Gärdenfors 2014). The focus in this chapter is on the descriptions of the sensory perceptions. They differ from object concepts such as WINE simply because they are not objects but sensations. VISION in this chapter is primarily treated as mapping on to the COLOR domain, but the link between conceptual space and SMELL, TASTE and TOUCH respectively is less straightforward.

The LOC framework, shown in Table 3.1, comprises a system of pre-meaning structures and a number of Construals, whose task it is to generate the profilings of the conceptual structures at the time of use in human communication. LOC thus assumes a system of both Ontological (conceptual) structures and Cognitive processes (Construals). Two types of conceptual structure are distinguished, namely Contentful (i.e. what the meanings are, e.g. ARTIFACT, ACTIVITY, COLOR) and Configurational structures (e.g. PART-WHOLE, SCALE, i.e. how the Contentful structures may be formatted by the Construals). The Construals form the dynamic part of the model, operating on the conceptual structures at the time of use. Concrete examples of how this works are presented in the subsequent chapters. While being firmly based in the Cognitive Linguistics framework, LOC also differs from the received view in two important respects. One is the explicit distinction between conceptual Configurations and Construals, which is not recognized in most Cognitive Linguistics treatments, the other is the view that words do not have meanings. Words are associated with a use potential that has been developed through encounters with language. When words are used in communication, they evoke specific meanings in the contexts where they are used. (For more details on this see Paradis 2005).

Table 3.1 Ontologies and cognitive processes in meaning construction

Ontologies (conceptual structures)		Cognitive processes
Contentful structures	Configurational structures	Construals
Pre-meanings relating to concrete spatial matters, to temporal events, processes and states, e.g. COLOR, SMELL, TASTE, TOUCH, WINE, GRAPE	*Pre-meanings of an image-schematic type which combine with the contentful structures*, e.g. SCALE, CONTRAST, BOUNDARY, PART-WHOLE	*Operations acting on the pre-meanings at the time of use,* e.g. Gestalt formation, Salience (e.g. metonymization), Comparison (e.g. metaphorization)

Adapted from Paradis (2005)

On the occasion of use in speech or writing, all language elements evoke the relevant parts of their meaning potentials, combining Contentful and Configurational structures through Construal operations. In the descriptions of sensations in wine reviews, the Ontological (conceptual) structures are spaces related to VISION, SMELL, TASTE and TOUCH, as exemplified in Table 3.1.

Depending on the role of the descriptor in the text, the Configurational structures in wine description may be structures such as SCALE, CONTRAST, BOUNDEDNESS, PART-WHOLE. These Configurations are viewing arrangements that are general and combinable with most Contentful meaning structures, if not all. The Construal mechanisms are responsible for the dynamics and the profiling of the linguistic expressions when they are used in discourse. Configurations are structuring elements that need the Contentful meaning structures to make sense. They are very few in comparison to the countless Contentful structures. In combination with Contentful structures in language use they are always "secondary", and do not have any status in the absence of their combining with Contentful domains, much like elements such as tense, definiteness, grading, aspect etc. As already mentioned, the final profiling of the meaning of a lexical item in human communication in discourse is carried out by the system of Construals. They operate on the conceptual structure at the time of use, in which case the profiling of a specific part of the whole meaning potential of, say, *wine* is brought about through a Construal of focus and salience as in metonymizations and/or through a Construal of Comparison as in contrasts, similes and metaphorizations.[2]

Also, in line with the broad framework of Cognitive Linguistics, Gärdenfors (2014), in his book on the *Geometry of Meaning*, highlights the importance of perception, in particular vision, for semantic representations. A central idea in his book is that the meanings can be described as organized abstract spatial structures, expressed primarily in terms of *dimensions*, *distances*, and *regions.* The foundational assumptions of Gärdenfors' framework are similar to those of LOC (2005); conceptual spaces are taken to be built up of quality dimensions. Dimensions may be separate, as is the case for, for example 'long', where LENGTH is the Contentful dimension,[3] while in other cases dimensions come in bundles. For instance, SPACE involves the dimensions of HEIGHT, WIDTH, and DEPTH, and COLOR the dimensions of HUE, SATURATION, and BRIGHTNESS. Gärdenfors refers to these spaces as domains. I prefer to refer to them as concepts, reserving the notion of domain for relational circumstances, i.e. when a concept serves in the background matrix, i.e. in the domain matrix of another concept, much like Langacker (1987) does.[4] Furthermore, Gärdenfors makes a point of the fact that

[2]The scope of this chapter does not allow for a discussion of Construals of Comparison, such as similes and metaphorizations. For treatments of that see Paradis and Eeg-Olofsson (2013) and Paradis and Hommerberg (in press).

[3]It should be noted that *long* may also evoke positive or negative evaluation (Paradis et al. 2012).

[4]A domain is a context for the characterization of a semantic unit. Domains are mental experiences, representational spaces, concepts and concept complexes. There are basic domains and abstract

topology and geometry allow us to talk about *nearness* and *distance* in conceptual space, i.e. if point *x* is nearer point *y* than point *z*, then *x is more similar* to *y* than to *z*. This is highly interesting for a range of different semantic phenomena, in particular for the phenomenon of polysemy and metonymy as noted by Cruse (2002) and Paradis (2004, 2011), for synonymy and antonymy (Paradis et al. 2009; Paradis and Willners 2011; Jones et al. 2012). Distance is an important concept in the characterization of cross-modal uses of words in this chapter. LOC has adopted Gärdenfors (2000) characterization of concepts as bundles of properties that are separate but correlated with one another. For instance, Gärdenfors argues that the concept of apple involves a very strong correlation between sweetness in the taste domain and the sugar content in the nutrition domain, but a weaker relation with the color red and sweetness. Properties are special cases of concepts in that they are based in a single domain, whereas concepts are based on more than one domain. LENGTH is a good example based on one quality dimension, but like other meanings also on Configurational structures, i.e. SCALE (Paradis 2001).

3.4 Analytical Systems for Wine Descriptions

It has now been established that wine reviewers' descriptions of the tasting event follow the journey of the wine from the glass through the nose and the mouth and finally into the gullet and/or the spittoon. The important task for the reviewers is then to transform the sensations of the wine into conceptual representations through language so that the sensations evoked in the tasting session become interpretable for the reader at the same time as the descriptions should arouse the reader's sensorium. Wine descriptions may be analytic or synthetic. The difference between those two is that, while the point of departure of analytic descriptions is the parts, the departure for synthetic descriptions is the whole or as Herdenstam (2004: 65–80) puts it: "[t]he analytical approach attempts to account for the sensory experience of wine, while the synthetic approach attempts to describe the total complexity of the whole", as already pointed out in the description of wine review (1) in Sect. 3.2.

This section presents the main types of recontextualization strategies for the description of the sensory perceptions in two different schemas of analytic terminologies, one using objects as descriptors and one using properties along scales. An example of the former system is a version of the Aroma Wheel (Noble et al. 1984) developed by The German Wine Institute,[5] and the other one is a schema

domains. Basic domains cannot be reduced to more fundamental but interrelated structures. Basic domains are primitive representational spaces such as TIME, SPACE, VISUAL SENSATIONS (COLOR), AUDITORY SENSATIONS (PITCH), TOUCH (TEMPERATURE, PRESSURE, PAIN), TASTE/SMELL. Langacker (1987:147–150) notes that all human conceptualization is presumably grounded in basic domains, mediated by chains of intermediate concepts. Any other concept or conceptual complex that functions as a domain is referred to as non-basic, or abstract.

[5]For information, see www.deutscheweine.com.

Table 3.2 A systematic approach to wine tasting according to the WSET

Wine and Spirit Education Trust (WSET)	
APPEARANCE	
clarity	bright – clear – dull – hazy – cloudy
intensity	
white	water-white – pale – medium – deep
rosé	pale – medium – deep
red	pale – medium – deep – opaque
color	
white	green – lemon – straw – gold – amber – brown
rosé	pink – salmon – orange – onion skin
red	purple – ruby – garnet – mahogany – tawny
other observations	Legs, bubbles, rim, color vs. core, deposits, etc.
NOSE	
condition	clean – unclean
intensity	weak – medium – pronounced
development	youthful – grape aromas – aged bouquet (*tired* – *oxidised*)
fruit character	fruity, floral, vegetal, spicy, woods, smoky, animal
	fermentation, aromas, ripeness, faults
PALATE	
sweetness	dry – off-dry – medium dry – medium sweet – sweet – luscious
acidity	flabby – low – balanced – sharp
tannin	astringent – hard – balanced – soft
body	thin – light – medium – full – heavy
fruit intensity	weak – medium – pronounced
alcohol	light – medium – high
length	short – medium – long

Adapted from Herdenstam (2004: 131)

of descriptors across Appearance (VISION), Nose (SMELL) and Palate (TASTE & TOUCH) organized along scales, as in Table 3.2, developed by the Wine and Spirit Education Trust (WSET). The existence of the Aroma Wheel does not make the WSET type of schema redundant and not vice versa either. On the contrary, both schemas can be seen as complementary methodologies and analytical systems that can be used as guiding tools.

The Aroma Wheel, which initially was developed by oenologists at the University of California at Davis for descriptions of smell, is a famous terminological attempt at a consistent and clear descriptor system (Noble et al. 1984). In the 30 years that have passed, the Aroma Wheel has been further developed in several different ways outside wine industry, e.g. the fragrance wheel for perfume industry, and new wheels for both whites and reds by the German Wine Institute with hints to taste as well. Figure 3.1 shows the German Aroma Wheel for red wines.

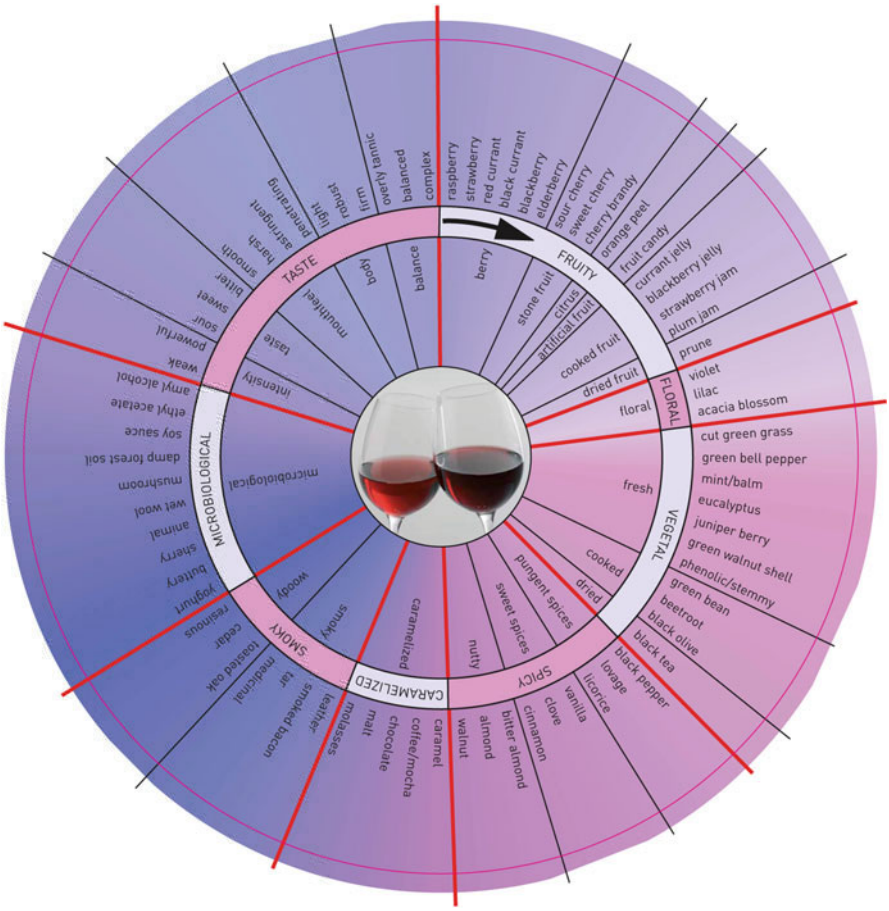

Fig. 3.1 The German Aroma Wheel for red wines (I am grateful to the German Wine Institute (http://www.deutscheweine.de/) for letting me use their picture)

As Fig. 3.1 shows, the descriptors of smell are organized into three tiers with the more general tiers close to the core and the more specific ones on the outskirts. In between are the category type labels. The most general tier contains property descriptors such as *fruity*, *chemical*, *spicy*, *earthy*, while the more specific ones are mostly contentful meanings referring to concrete objects such as *blackberry*, *fresh bread*, *oak*, and *cinnamon*. The tiers are connected from the core and outwards in a hierarchical system where, for instance, *fruity* subdivides into *citrus*, *berry*, *tree fruit*, *melon, tropical fruit*, *cooked fruit*, *artificial fruit*, which in turn subdivides into *orange*, *grapefruit*, *lemon*, *lime* for CITRUS and *blackberry*, *raspberry*, *strawberry*, *black currant* for BERRY, and so on.

In contrast to the type of terminological system represented by aroma wheels, whose main focus is on smell, the WSET approach to wine description covers

property descriptors in the domains of VISION, SMELL, TASTE and TOUCH, referred to as APPEARANCE, NOSE and PALATE (conflating TASTE and TOUCH). Each of these perceptual domains comprises a number of dimensions which are expressive of a certain perceptual property of WINE in the corresponding parts of the experiential procedure. The characterization of the wine is done on the basis of the identifications of the different Contentful dimensions of the different sensory perceptions. Table 3.2 shows that most of the descriptors are properties organized along scales of opposition, with the exception of the smell descriptors, under NOSE, which include properties of worldly objects, such as *fruity, floral, smoky* and *animal*, i.e. the sensations these objects produce, although construed as scalable dimensions. The visual dimensions comprise information about CLARITY, along a scale from *bright* to *cloudy*, valenced from positive to negative, INTENSITY, basically going from *pale* to *deep*, and COLOR, divided into the traditional types, i.e. WHITE, ranging from *water-white* to *deep*, ROSÉ, from *pale* to *deep*, and RED, from *pale* to *opaque*. In addition to those quality dimensions, a list of other visual observations is offered in terms of object categorization, i.e. *legs* and *bubbles.*

Next, the olfactory terminology involves CONDITION, i.e. *clean* vs. *unclean*, INTENSITY, from *weak* to *pronounced*, DEVELOPMENT, from *youthful* to *tired,* and a list of object-related terms, such as the ones in the Aroma Wheel above. Finally, the gustatory and tactile observations are based on the dimensions of SWEETNESS, from *dry* to *luscious,* ACIDITY, from *flabby* to *sharp*, TANNIN, from *astringent* to *soft*, BODY, from *thin* to *heavy*, FRUIT INTENSITY, from *weak* to *pronounced*, ALCOHOL, from *light* to *high* and finally, LENGTH from *short* to l*ong*. The properties are located in different ranges of the SCALE Configurations. Some of the dimensions are more closely correlated with one another such as in the case of Gärdenfors' example of the characterization of 'apple', in which case the color of red apples is correlated to sweetness, while the size of the apple is not.

In the framework of LOC, the dimensions of the sensory perceptions are all Contentful dimensions with different properties of those dimensions along SCALE Configurations, see Table 3.1. The properties at the opposite ends of the scales are antonyms. For the interpretation of antonyms speaker have to make use of a Construal of Comparison, i.e. when we say that a wine is *dry* rather than *luscious* we are in effect comparing their SWEETNESS. In order to evoke this constrasting Gestalt, we construe dimensionally aligned Comparisons (Paradis and Willners 2011). The Construal of the antonymic scale structure and its conceptual structures are shown in Table 3.3.

Table 3.3 Antonymy in LOC

Ontologies (conceptual structures)		Cognitive processes
Contentful structures	Configurational structures	Construals
DIMENSION (x)	SCALE, BOUNDEDNESS, CONTRAST	Gestalt: dimensional alignment
		Comparison

From the point of view of antonymy as a Construal in human communication, antonyms in terminologies such as this wine terminology are similar to antonyms in natural language in that they are Construals of binary contrasting elements meant to be opposites. However, a terminology, like the one in Table 3.2, is consciously structured by scientists, and the descriptors are defined and specified in relation to the Contentful structure of a DIMENSION (x), configured as a SCALE of CONTRAST in a dimensional alignment Gestalt formed by Comparison. This state of affairs is essentially the same in natural language, with the difference that, in everyday language use, form-meaning pairings are not defined by individuals, rather, they evolve and emerge in speech communities. In this respect, antonymy in terminologies is essentially the opposite of antonymy in natural language.

3.5 Descriptors of Sensory Experience

This section investigates the main recontextualization strategies for the description of the sensory perceptions in our database. The wine reviews in our database are mainly what Herdenstam (2004: 65–80) refers to as analytical descriptions, i.e. the sensory perceptions are described separately from one another by means of terminologies that are designed to facilitate the description and the interpretation of the perceptive experience. However, as already mentioned, nearly all the reviews also include more holistic or synthetic comments that attempt to describe the complexity of the whole experience. As will become evident, the descriptions in the wine reviews are rendered through expressions of properties of the sensory modalities and properties of objects with a high degree of cross-modal overlap.

3.5.1 Properties and Objects

This section starts with a presentation of examples of commonly used descriptors in the wine database. As indicated at the beginning, the descriptions of the sensory modalities are normally presented in the order from vision, smell, through taste and touch. The latter two are often conflated. The reason for the conflation of taste and touch is that they are often very difficult to tease apart. Experientially, they are two sides of the same coin. Putting something in our mouths necessarily gives rise to a feeling of its texture. The domain of SMELL attracts most descriptor types, closely followed by TASTE/TOUCH. VISION attracts the fewest. Based on a search of the database using premodifiying descriptors of seed words such as *color, aroma/s, nose, scent/smell, flavor/s, taste, body, palate* and *texture*, the proportions of the number of descriptors are 50 % for SMELL and 41 % for TASTE/TOUCH and 9 % for VISION. For a more detailed discussion of this, see Paradis and Eeg-Olofsson (2013). Table 3.4 gives an overview of the descriptors across VISION, SMELL, and TASTE/TOUCH.

Table 3.4 List of examples of different descriptors of VISION, SMELL, and TASTE/TOUCH

VISION	SMELL	TASTE/TOUCH
dark, light, deep, soft, solid, shallow, bright, dense, brilliant, full, strong, weak, young, thick, ...	*deep, thin, tight, full, weak, huge, focused, expansive, ...*	*big, chewy, dense, dry, deep, fat, pure, rich, ripe, supple, sweet, long, austere, ...*
black, blue, amber, crimson, garnet, deep-ruby, green, purple, plum, red, white ...	*apricot, earthy, floral, game-like, oaky, Oriental, musty, spice-box, perfumed, almond, apple, blackberry, rose, nut, peach, ...*	*textured, creamy-textured, silken-textured, concentrated, multi-dimensional, sustained, oily, ...*
	animal-like, caramel-infused, chocolate-drenched, cassis-scented, ...	

As indicated by the examples in Table 3.4, some of the descriptors of visual experiences are expressed through lexical items that are common core expressions in the domain of sight in language more generally, *light, dark, brilliant*, and through other dimensional properties such as *deep, soft, strong, thick* and *young*, while some others are more specific and also more clearly spring from names of objects (*ruby, straw, gold*) but are conventionalized color words in common parlance. The former are all gradable scalar dimensions, while the latter are like object concepts in that they are defined through a set of quality dimensions and not through a range along a SCALE. The descriptors of SMELL may also be described through general dimensional property words such as *weak*, *deep, thin, full*, but it differs from both VISION and TASTE/TOUCH in that it is mainly described through derivations of terms referring to objects, such as *fruity, floral, spicy,* and *smoky*, and the objects themselves, e.g. *apricot, spice-box* and *blackberry*. The descriptors of TASTE/TOUCH are mainly expressed by both properties along general spatial dimensions, such as *big, deep, long,* and more specific property words along a SCALE dimensions, such as *chewy, supple, austere*, *textured* and *oily*.

3.5.2 Cross-Sensory Descriptors and Their Meanings

As shown in the previous sections, property descriptors differ across the modalities, but there are also many descriptors that are the same across two or more modalities, e.g. *deep, soft, big, bright, light, thick, thin, solid, strong, shallow, sweet, smooth, round, tight, sharp, dense, warm, weak, dark, broad, bright, fat* and *hard.* They are frequent property terms of salient spatial dimensions, important for the scaffolding of the lexical semantic structure of language and cognition in general and sensory cognition in particular (Paradis et al. 2009; Paradis and Eeg-Olofsson 2013). *Deep*, for instance, is a qualifier that is used across all four modalities. In addition to

expressions such as *deep nuttiness, deep garnet, deep raspberry, deep mouth-coating*, *deep* also qualifies the sensory perceptions directly *deep colors, deep scents, deep aromas, deep texture, deep flavors,* and *deep finish*. These cross-overs are true of most of the property words. For instance, color descriptors such as *black* and *white* are commonly used as modifiers in the descriptions of object descriptors for SMELL.[6] For instance, *black fruits, black cherries, black chocolate, black raspberries, black currants,* and *white flowers, white peaches, white pepper, white fruit, white currants*. Interestingly, there are also a couple of cases where *white* directly modifies *aroma* without a specification of an object, as in (2).

(2) The 2001 Chardonnay Marina Cvetic, in addition to its ripe lemon and **white aromas** and subtle oak spices, manages to combine a tonic acidity to the volume and viscosity of the flavors.

In the literature, property words have been treated as synesthetic metaphorical extensions of one literal meaning (Shen 1997; Shen and Gadir 2009), when applied to say *soft*, would involve an extension from TOUCH to VISION, SMELL and TASTE, and for combinations, such as *soft colors, soft nose, soft flavors, soft mouth-feel, soft finish*, the argument would be that *soft mouth-feel* is the only congruent, literal meaning all the others are metaphorical extensions from the domain of TOUCH. According to that approach, the descriptions of perceptions are characterized by synesthesia from lower to higher modalities. In his work on synesthesia in poetry, Ullman (1945) proposes a hierarchy and a directional principle of sensory perceptions in metaphors, i.e. from TOUCH > TASTE > SMELL to SOUND and VISION. His proposal has been acknowledged and further developed in different areas of research by a number of scholars, e.g. Williams (1976), Lehrer (1978), Viberg (1984), Sweetser (1990), Shen (1997), Popova (2003, 2005), Plümacher and Holz (2007), Shen and Gadir (2009). On the basis of Ullman's hierarchy and directional principle, Shen (1997) and Shen and Gadir (2009) formulate the *Conceptual Preference Principle* according to which the preferred direction of mappings in what they refer to as synesthetic metaphorization is from the lower sensory domains of touch and taste, both of which require direct contact with the perceiver, to the higher modalities of vision and sound, which do not require direct contact with the perceiver (see Traugott and Dasher 2005: 72 Figure 2.4).

The *Conceptual Preference Principle* entails two things, (i) that meanings do not extend downwards from say VISION to SMELL, and, (ii) that the extended or metaphorical senses are different from the source sense, i.e. the expressions are polysemous. The data used in this study challenge the *Conceptual Preference Principle*. They do not confirm the conceptual preference pattern described above, and therefore cast doubts on the grounds for their claims for polysemy and metaphor. As a case in point we may take the cross-sensorial patterning of *soft* and *dark*.

[6]Please, note that this does not only apply to monochromatic but also to chromatic descriptions, see Fig. 3.1.

According to the received view described above, it is assumed that when *soft* is used about smell, taste or color the meanings extend from TOUCH to these other sensorial domains, which in effect means that *soft* in these different uses is polysemous and metaphorical in all of them, but the touch sense. If we test this, using traditional semantic tests of co-ordination, it becomes clear that no strong zeugmatic effect is created. For instance, *both the aroma and the color are soft*, *both the color and the flavors are soft*, *both the mouth-feel and the color are soft*. Hence, their meanings are not autonomous; they are not different senses of *soft.* In contrast to *soft*, the alleged source meaning of which is at the beginning of the hierarchical system (TOUCH), the source meaning of *dark* is the end-point (VISION). This poses severe problem for the directional principle because the target meaning goes in the wrong direction, i.e. from VISION to SMELL and TASTE instead of the other way round. Like *soft, dark* is also cross-sensorial. In addition to *dark colors*, we find *dark aromas, dark flavors,* and all kinds of object descriptors such as *dark plum,* and *dark tobacco*. There is no evidence in favor of a polysemy analysis of such uses. Again, traditional semantic truth tests using syntagmatic constraints do not give rise to zeugmatic interpretations. Sentences such as *both the aroma and the color are dark*, or *both the color and the flavors are dark* do not give rise to aberrant zeugmatic readings or puns. What is obvious here, unlike when property words like *soft* and *dark* serve as qualifiers of sensory perceptions directly, is that they do not seem to be autonomous meanings, but in combinations with entities of different kind such as in ?*both the aroma and the sky are dark*, or ?*both the flavor and the sofa are soft,* they are autonomous and cannot be combined. This raises the question of whether property words such as *dark* or *soft* have two senses when they are used to qualify sensory perceptions the way they are here. Judging from the outcome of the zeugma test, this does not seem to be the case. This takes us to the next aspect of this discussion, which concerns whether the cross-sensual uses of descriptors involve metaphorization.

According to the definition of metaphor in LOC, which is also the received definition of metaphor in Cognitive Linguistics, metaphorization is a construal of Comparison across different domains with invariable configurational structuring (Lakoff 1987; Paradis 2005). In the case of our descriptors, it is not clear how and what aspects of meanings are compared across the sensory domains, when they are used cross-sensually. When *soft* extends from TOUCH to SMELL or VISION, it is still the soothing sensation that is at stake cross-sensually. Granted that metaphor is defined as a mapping across domains where the Configurational structure is kept constant, it is hard to see what the Comparisons across Contentful domains would be and what the invariant Configuration would be for expressions that involve sensory perceptions such as *dark aromas, dark colors, dark flavors*. However, if we instead imagine contexts such as *dark personality* or *dark story,* the metaphorical cross-over from VISION to PERSONALITY and STORY involves a Comparison across domains where darkness is associated with danger or sadness. Both are negative in contrast to its opposite LIGHT, which is positive. The contrastive valence is thus the invariant Configuration, much like in the ancient Chinese philosophy where Yin and Yang represent negativity and positivity respectively and where the literal meanings in

actual fact are 'dark' and 'light' (Osgood and Richards 1973). In her work on the distinction between literal and metaphorical meanings, Rakova (2003: 49, 142) notes that perception of cross-modal similarities is universal, systematic and present in early childhood. She points out that we may think of concepts such as BRIGHT, SHARP and COLD as primitive concepts spanning all domains of sensory experience, and they are better thought of as neural configurations responsive to certain stimuli. *Why* some words came to be regarded as more accessible or more primitive has not yet received a convincing explanation. An important reason may be that some experiences are more important than others in our daily lives in a given situation. This said, a note of caution is in place: Anthropologists and language typologists repeatedly point out that the differences across cultures may be greater than we think due to a paucity of research on these things in cultures other than Western cultures (e.g. Howes 2003, 2013; Majid and Levinson 2011; Caballero and Díaz Vera 2013; Majid and Burenhult 2014; Caballero and Paradis 2015). This means that rather than metaphorization, which involves Comparison across domains, cross-sensual uses are better thought of as transitions across primary domains, which do not involve Comparison. Such transitions across primary domains in human language are thus monosemous and syncretic rather than metaphorical and polysemous.[7]

The question then is *why* it is that no zeugmatic readings are created for combinations of sensory perceptions (*both the aroma and the color are dark*), but for combinations of different objects (?*both the flavor and the sofa are dark*) or abstract entities *(?both the flavor and the story are dark*). If we accept that properties of sensory information do not extend from a source but instead receive their interpretations on the same conditions in the various different sensory domains, the analysis is one in favor of a monosemy approach instead of a polysemy approach. The reason for monosemy in language is due to the conceptual nearness of the sensory representations of the experiences, as opposed to the conceptual distance of say FLAVOR and PEOPLE, or FLAVOR and STORY. My proposal thus appeals to Gärdenfors' (2014) topological notion of *distance* described above and to previous treatments of the continuum from polysemy to monosemy as a reflection of distance in conceptual space (Cruse 2002; Paradis 2004, 2005, 2008, 2011). Cruse (2002) describes the total meaning of a linguistic element as a pattern of readings in conceptual space. He describes readings as bounded regions in conceptual space, which tend to cluster in groups, and as such they show different degrees of salience and cohesiveness. Between these groups of readings there are regions that are relatively sparsely inhabited. They are sense boundaries and sense distinctions and polysemy are considered to be a function of distance and boundaries in conceptual space is consistent with LOC and already developed for treatments of language change (Paradis 2011). This approach to the modelling of meaning differences

[7]For a similar argument against a metaphor/polysemy account of cross-modal sensory word meanings, see Johnson (1999). In a study of the acquisition of *see* he argues for a (first acquired) general meaning of *see* for both vision and understanding, rather than the metaphoric extension of vision to cognition and knowledge.

also means that the notion of a sense boundary and boundaries between readings within a sense are closely related to the degree of autonomy of the clusters that the boundaries delimit. Senses exhibit strong signs of autonomy and they are kept apart by substantial boundaries, whereas readings within a sense are only weakly autonomous or not autonomous at all and separated by less than substantial boundaries. It is the symptoms of autonomy that are highlighted through various definitional tests such as the zeugma test that provide the evidence for boundaries. The different uses of say *soft* and *dark* are not as distinct as different senses but are just readings of close conceptual representations of sensory meanings.

This reasoning does not only apply to properties of the sensory perceptions but also to the activation of properties of object concepts used to describe the sensory perceptions. For instance, *blackberry, apple, lemon, vanilla, cedar, chocolate* and *tobacco* all evoke the conceptual structures of their meaning potential, i.e. BLACKBERRY, APPLE, LEMON, VANILLA, CEDAR, CHOCOLATE and TOBACCO. In its discursive context in wine reviews, the descriptor *blackberry* is used to evoke the smell of a wine. Through the use of a dark object we know that the wine described is a red wine and the taste of such a wine is likely to be rich and opulent. The quality dimensions of the descriptors are thus strongly correlated. Although used about smell, this is how the other properties of the object descriptor range over vision, taste and touch as well. This does not mean that the meanings of object descriptors are polysemous. Like the uses of adjectives such as *soft* and *dark*, the readings of the nominal descriptors are monosemous.

This closeness of the sensory knowledge domains has been shown both through textual studies and experimentally. As already pointed out, the main strategy of describing SMELL is through the use of objects, as in (3).

(3) A blockbuster effort, the 2005 boasts an inky/blue/purple color along with aromas of crème de cassis, blackberries, truffles, fruitcake, and toasty oak. Pure and full-bodied with significant extract, tannin, acidity, and alcohol, this stunning wine should be very long lived.

The Construal of the meaning of the smell is through the smell of *crème de cassis, blackberries, truffles, fruitcake, and toasty oak*. They are construed with focus of attention on smell as the salient dimension through a WHOLE FOR PART Construal. The concrete objects are used to evoke contingent properties that the objects produce. This is a Construal of metonymization, which in the case of wine descriptions does not give rise to multiple meanings of the object descriptors but the activation of a zone within a concept, i.e. within monosemy (for a detailed description of the differences, see Paradis 2004, 2011).

Even though reference to objects such as the ones above is mainly used to describe olfactory characteristics, it is important to note that these objects also provide visual, gustatory and tactile information in the domain matrix. In spite of the fact that they are not highlighted when they are used about SMELL, they form the base of the profiled olfactory information. The use of objects for identification of SMELL is motivated by the fact that smelling is made possible through a source, and hence we represent and understand SMELL through these sources. Also, concrete

word meanings, in contrast to abstract ones, elicit qualitatively different processing in the form of mental images in that they evoke rich sensory experiences which are intimately tied up with our experiences in life (Huang et al. 2010).

Also, it should be mentioned that the importance of the visual properties of the object descriptors has been found to be of crucial importance for the aesthetic expectations of SMELL, TASTE and TOUCH. Even though the visual descriptors are fewer in the reviews, they are not less important. On the contrary, we drink with our eyes first. It has been shown in wine tasting sessions among professionals that visual stimuli are capable of hi-jacking other sensual perceptions. Morrot et al. (2001) show that even professional wine tasters may be fooled by the color of the wine, starting to describe white wines dyed red, as if they were red. On the basis of their psychophysical experiment in which the smell of a white wine artificially colored red with an odorless dye was described by means of descriptors used about red wines, Morrot et al. (2001) propose that the existence of this synesthesia of smell and vision in wine description is psychologically grounded. The consistency of color-related descriptions is confirmed by the descriptions of the wines in our database. The large number of wine reviews allows us to be able to establish that there are clear differences between smell descriptors of red wines and of white wines. As shown in Table 3.5, red wines are mainly described by dark object, while the opposite is true of white wines.

Red wines are mostly described through "darkish" objects, such as licorice, blackberry, tar and chocolate, while white wines are mostly rendered through light-colored objects, such as honey, peach, melon and grapefruit. Some of the descriptors for reds and whites are the same. *Spice* is one of those. However, as one descriptor among several others in descriptions, the actual spices referred to differ. This highlights the importance of the correlations of dimensions in the creation of meaning. Consider the contexts for *spices* for a red and a white wine in (4) and (5), respectively.

(4) It possesses enthralling aromas of black raspberries, dark cherries, beef blood, and Asian **spices** that give way to an oily-textured, magnificently concentrated, highly-refined, and very focused personality.
(5) This decadent offering is studded with lychees, yellow plums, roses, assorted white flowers, and **spices** whose effects linger in its extensive finish.

Table 3.5 Common object descriptors for reds and whites: dark objects and light objects respectively

Red wines	White wines
Cassis, spice, cherry, currant, licorice, blackberry raspberries, mineral, black-cherry, chocolate, plum, pepper, blueberry, wood, oak, tar ...	*Apple, pear, peach, flower, honey, oil, sugar, butter, orange, herb, spice, honeysuckle, pineapple, melon, vanilla, apricot, grapefruit, almond, hazelnut, salt ...*

In (4), *spices* in the description of the red wine is surrounded by dark objects, *black raspberries, dark cherries, beef blood*, which is not the case in the description of the white wine (5) where *spices* is surrounded by *lychees, yellow plums, roses, assorted white flowers*, i.e. light-colored objects.

Summing up, I propose that the use of object descriptors for smell, spilling over into vision, taste and touch, is grounded in very weak autonomy at the conceptual level, or what Morrot et al. (2001) refer to as 'synesthesia' of sensory information, and the lexical syncretism of property expressions is evidence of conceptual nearness within monosemy. If you taste something you also smell it and feel it, and if you see something you also have an idea of its smell and taste (even though the actual smelling, tasting and feeling cannot be experienced in the absence of the object). In other words, the conceptual structures of sensory meanings of the different perceptions are not autonomous. This paves the way for syncretism at the lexical level. The impact of color for the other modalities is very strong and the absence of words for smell and the ontological cross-over of sensory modalities are taken to be symptoms of *real* synesthesia in the wine tasting event by Morrot et al. (2001). Yet, in spite of the sensory power of vision as a point of departure for the experience, expressions of vision do not dominate the descriptions in the reviews and the sensory importance of appreciation of the wine drinking event as such.

3.6 Conclusion

This chapter is concerned with how experiences of sensory stimuli of VISION, SMELL, TASTE, and TOUCH are recontextualized and rendered into language. The data used are terminological schemas of descriptors used by professional wine critics and actual reviews of individual wines in which critics translate their experiences in the tasting practice into written discourse. The focus has been on the types of conceptual structures used in the descriptions across the sensory modalities, both in terms of Content and Configurations, and how they are construed in the discourse. Observations made on the basis of schemas, which are constructed by professionals, and the more journalistic translations of sensory experience into written discourse by wine critics are explicated in the framework of LOC and Gärdenfors' geometrical notion of distance in conceptual space.

It has been shown that the visual appearance of the wine is mainly described by color terms, sometimes with the addition of properties of clarity and intensity as in *a dense ruby/purple color.* The gustatory and tactile experiences are also primarily described through properties along Contentful dimensions such as SWEETNESS (*dry, sweet*), TANNIN (*astringent, soft*) etc., while olfactory experiences, which make up the lion's share of the descriptions, make use of concrete objects, as in *sweet tobacco, black currants, leathery aromas*, where the focus of attention is on the smell of these objects, which mainly come from domains such as FRUIT, HERBS, SPICES, FLOWERS, PLANTS, SWEETS, BEVERAGES, MINERALS, BUILDINGS, FOOD, LIVING CREATURES. Linguistically, these descriptions are

construed through a process of zone activation, i.e. the zooming in on their smell as a reference point in the conceptual complex as a whole. This zoomed-in aspect of meaning is contingent and does not create a new meaning of the words, but an activation of a zone of a conceptual within their domain complex. Also, the color of the objects used as descriptors is important and differently colored objects are used to describe differently colored wines.

Another finding that emerges from the study is that not only are the object descriptors used across the sensory perceptions, but there are also a fair number of property descriptors that are used across two, three or all of the sensory domains. In the literature, this syncretism of property words such as *dark* and *soft* has previously been analyzed as cases of metaphorization and polysemy. This is an approach that is challenged in this chapter. The reasons are that the properties expressed by such descriptors are slightly different because they are instantiated in different domains, but the domains are very closely interrelated domains and therefore only give rise to reading differences rather than sense distinction. Using property words such as *dark* and *soft* cross-modally does not give rise to any zeugmatic readings, and it cannot be reasonably argued that we make Comparisons across the sensory perception with invariant configurational structures across the sensory representations. Instead, when words expressing properties along dimensions are used as modifiers of sensory perceptions they are monosemous. When such property words are used to qualify meanings that are not primary, i.e. that do not relate directly to sensory perceptions, such as distinct objects or abstract phenomena, sense distinctions are created because the modified concepts are not located closely to one another in conceptual space. They are autonomous.

The sensory perceptions form bundles of the same concept complex and in the event of experience they cannot be separated. There is a saying "We eat with our eyes first", which indicates that visual experience cannot be separated form smelling, tasting and feeling (under normal circumstances). This is evidenced in the chapter by the difference of colors of descriptors for red wines (dark colors) and white wines (light colors) as well as evidence from experiments pointing to the deterministic influence of sight for smelling, tasting and feeling. At the conceptual level this closeness results in strongly interrelated sensory representations that are dependent on one another and monosemy and syncretism in language. These type of data are also a challenge to the Conceptual Preference Principle, since there do not seem to extensions from a single source domain into the other domains.

References

Barsalou, L. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645.

Beveridge, M. E. L., & Pickering, M. J. (2013). Perspective taking in language: Integrating the spatial and action domains. *Frontiers in Neuroscience, 7*, 277. doi:10.3389/fnhum.2013.00577.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15*(11), 527–536.

Caballero, R. (2009). Cutting across the senses. Imagery in winespeak and audiovisual promotion. In C. Forceville & E. Urios-Aparisi (Eds.), *Multimodal metaphors* (pp. 73–94). Berlin/New York: Mouton de Gruyter.

Caballero, R., & Díaz Vera, J. (Eds.). (2013). *The embodied soul – Explorations into human sentience – Imagination, (e)motion and perception*. Berlin: de Gruyter Mouton.

Caballero, R., & Paradis, C. (2013). Perceptual landscapes from the perspective of cultures and genres. In R. Caballero & J. Díaz Vera (Eds.), *The embodied soul – Explorations into human sentience – Imagination, (e)motion and perception* (pp. 77–105). Berlin: de Gruyter Mouton.

Caballero, R., & Paradis, C. (2015). Making sense of sensory perceptions across languages and cultures. *Functions of Language. 22*(1), 1–19

Classen, C., Howes, D., & Synnott, A. (1994). *Aroma. The cultural history of smell*. London/New York: Routledge.

Croft, W., & Cruse, A. (2004). *Cognitive lingusitics*. Cambridge: Cambridge University Press.

Cruse, A. (2002). The construal of sense boundaries. *Revue de Sémantique et Pragmatique, 12*, 101–119.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press.

Gärdenfors, P. (2014). *Geometry of meaning: Semantics based on conceptual spaces*. Cambridge, MA: MIT Press.

Gibbs, R. (1994). *The poetics of mind: Figurative thought, language and understanding*. New York: Cambridge University Press.

Gluck, M. (2003). Wine language. Useful idiom or idiot-speak? In J. Aitchinson & D. Lewis (Eds.), *New media language* (pp. 107–115). London: Routledge.

Herdenstam, A. (2004). *Sinnesupplevelsens estetik. Vinprovaren, i gränslandet mellan konsten och vetenskapen*. Stockholm: Stockholm Dialoger.

Hommerberg, C. (2011). *Persuasiveness in the discourse of wine: The rhetoric of Robert Parker*. Växjö: Linnaeus University Press.

Hommerberg, C., & Paradis, C. (2014). Constructing credibility through representations in the discourse of wine: Evidentiality, temporality and epistemic control. In D. Glynn & M. Sjölin (Eds.), *Subjectivity and epistemicity. Stance strategies in discourse and narration* (pp. 211–238).

Howes, D. (2003). *Sensual relations. Engaging the senses in culture and social theory*. Ann Arbor: The University of Michigan Press.

Howes, D. (2013). Postscript to Senuous Cognition: The language of the senses. In R. Caballero & J. Díaz Vera (Eds.), *Senuous cognition: Explorations into human sentience: Imagination, (E)motion, and perception* (pp. 293–299). Berlin: de Gruyter Mouton.

Huang, H. W., Lee, C. L., & Federmeier, K. D. (2010). Imagine that! ERPs provide evidence for distinct hemispheric contributions to the processing of concrete and abstract concepts. *NeuroImage, 49*, 1116–1123.

Johnson, C. (1999). Metaphor vs. conflation in acquisition of polysemy: The case of *see*. In M. K. Hiraga, C. Sinha, & S. Wilcox (Eds.), *Cultural, psychological and typological issues in Cognitive Linguistics*. Amsterdam: John Benjamins.

Jones, S., Murphy, M. L., Paradis, C., & Willners, C. (2012). *Antonyms in English: Construals, constructions and canonicity*. Cambridge: Cambridge University Press.

Kerren, A., Kyusakova, M., & Paradis, C. (2013). From culture to text to interactive visualization of wine reviews. In F. T. Marchese & E. Banissi (Eds.), *Knowledge visualization currents: From text to art to culture* (pp. 85–110). Heidelberg: Springer.

Lacey, S., Stilla, R., & Sathian, K. (2012). Metaphorically feeling: Comprehending textural metaphors activates somatosensory cortex. *Brain and Language, 120*(3), 416–421.

Lakoff, G. (1987). *Women, fire and dangerous things*. Chicago: The University of Chicago Press.

Langacker, R. (1987). *Foundations of cognitive grammar*. Stanford: Stanford University Press.

Langacker, R. (1999). *Grammar and conceptualization*. Berlin: Mouton de Gruyter.

Lehrer, A. (1978). Structures of the lexicon and transfer of meaning. *Lingua, 45*, 95–123.

Majid, A., & Burenhult, N. (2014). Odors are expressible in language, as long as you speak the right language. *Cognition, 130*, 266–270.

Majid, A., & Levinson, S. C. (2011). The senses in language and culture. *Senses and Society, 6*(1), 5–18.

Morrot, G., Brochet, F., & Dubourdieu, D. (2001). The color of odors. *Brain and Language, 79*, 309–320.

Noble, A., Arnold, R., Buechsenstein, J., Leach, E., Schmidt, J., & Stern, P. (1987). Modification of a standardized system of wine aroma terminology. *American Journal of Enology and Viticulture, 38*(2), 143–146.

Osgood, C. E., & Richards, M. M. (1973). From Yang and Yin to and or but. *Language, 49*(2), 380–412.

Paradis, C. (2001). Adjectives and boundedness. *Cognitive Linguistics, 12*(1), 47–66.

Paradis, C. (2004). Where does metonymy stop? Senses, facets and active zones. *Metaphor and Symbol, 19*(4), 245–264.

Paradis, C. (2005). Ontologies and construals in lexical semantics. *Axiomathes, 15*, 541–573.

Paradis, C. (2008). Configurations, construals and change: Expressions of degree. *English Language and Linguistics, 12*(2), 317–343.

Paradis, C. (2009). This beauty should drink well for 10–12 years: A note on recommendations as semantic middles. *Text & Talk, 29*(1), 53–73.

Paradis, C. (2011). Metonymization: Key mechanism in language change. In R. Benczes, A. Barcelona, & F. Ruiz de Mendoza Ibáñez (Eds.), *What is metonymy? An attempt at building a consensus view on the delimitation of the notion of metonymy in Cognitive Linguistics* (pp. 61–88). Amsterdam: John Benjamins.

Paradis, C. (in press). Meanings of words: Theory and application. In U. Hass & P. Storjohann (Eds.), *Handbuch Wort und Wortschatz* (Handbücher Sprachwissen-HSW Band 3). Berlin: Mouton de Gruyter.

Paradis, C., & Eeg-Olofsson, M. (2013). Describing sensory perceptions: The genre of wine reviews. *Metaphor & Symbol, 28*(1), 1–19.

Paradis, C., & Hommerberg, C. (in press). We drink with our eyes first: Multiple sensory perceptions and mixed imagery in wine reviews. In R. Gibbs (Ed). John Benjamins.

Paradis, C., & Willners, C. (2011). Antonymy: From conventionalization to meaning-making. *Review of Cognitive Linguistics, 9*(2), 367–391.

Paradis, C., Willners, C., & Jones, S. (2009). Good and bad opposites: Using textual and psycholinguistic techniques to measure antonym canonicity. *The Mental Lexicon, 4*(3), 380–429.

Paradis, C., van de Weijer, J., Willners, C., & Lindgren, M. (2012). Evaluative polarity of antonyms. *Lingue e Linguaggio, 2*, 199–214.

Plümacher, M., & Holz, P. (Eds.). (2007). *Speaking of colors and odors*. Amsterdam: John Benjamins.

Popova, Y. (2003). 'The fool sees with his nose': Metaphorical mappings in the sense of smell in Patrick Süskind's Perfume. *Language and Literature, 12*, 135–151.

Popova, Y. (2005). Image schemas and verbal synaesthesia. In B. Hampe (Ed.), *From perception to meaning: Image schemas in cognitive linguistics* (pp. 1–26). Berlin/New York: Mouton de Gruyter.

Rakova, M. (2003). *The extent of the literal: Metaphor, polysemy and theories of concepts*. New York: Palgrave Macmillan.

Shen, Y. (1997). Cognitive constraints on poetic figures. *Cognitive Linguistics, 8*(1), 33–71.

Shen, Y., & Gadir, O. (2009). Target and source Assignment in Synaesthetic Possessive Constructions. *Journal of Pragmatics, 41*(2), 357–371.

Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.

Talmy, L. (2000). *Toward a cognitive semantics*. Cambridge, MA: MIT Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Traugott, E. C., & Dasher, R. B. (2005). *Regularity in semantic change* (Cambridge studies in linguistics). Cambridge: Cambridge University Press.
Ullman, S. (1945). Romanticism and synaesthesia: A comparative study of sense transfer in Keats and Byron. *PMLA, 60*(3), 811–827.
Viberg, Å. (1984). The verbs of perception: A typological study. *Linguistics, 21*(1), 123–162.
Williams, J. (1976). Synaesthetic adjectives: A possible law of semantic change. *Language, 52*, 461–478.
Wittgenstein, L. (1977). *Remarks on colour*. Oxford: Basil Blackwell.
Zucco, G. M. (2007). The unique nature of a memory system. In M. Plümacher & P. Holz (Eds.), *Speaking of colors and odors* (pp. 155–165). Amsterdam: John Benjamins.

Chapter 4
Conceptual Spaces, Features, and Word Meanings: The Case of Dutch Shirts

Joost Zwarts

Abstract This paper explores how a conceptual space for the representation of word meanings can be constructed and visualized for one particular domain, namely Dutch words for different types of shirts. It draws on earlier empirical corpus-based research that has identified different features for uniquely describing each of these types and different ways in which they are lexically described in fashion magazines. The present study defines a metric that makes it possible to construct a feature-based space in which the extension of each of the Dutch shirt terms can be visualized and in which it is possible to study the distribution of words and the validity of different constraints on that distribution: conjunctivity, convexity, connectivity, coherence, and centrality. Although the paper concludes that definite conclusions about these constraints are only possible on the basis of more complete lexical datasets, it demonstrates the potential of the conceptual space approach for studying word meanings.

4.1 Conceptual Spaces and Semantic Maps

One way of doing lexical semantics is by studying particular meaning domains as *conceptual spaces* or *semantic maps* (see Gärdenfors 2000; Haspelmath 2003, respectively, for general overviews). The idea is that a domain consists of a set of values as points, geometrically structured in a particular way, with lexical categories (extensions) as regions. The geometrical structure of the domain could be assumed to be universal, but languages divide it up in different ways. One well-known example is the color space, with its dimensions of hue, saturation, and brightness (Kay et al. 2009). Another example is the graph of functions of indefinite pronouns (Haspelmath 1997), as shown in Fig. 4.1.

Part of this paper was presented at the workshop *Conceptual Spaces at Work*, Lund, May 24–26, 2012.

J. Zwarts (✉)
Department of Languages, Literature and Communication, Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands
e-mail: J.Zwarts@uu.nl

F. Zenker, P. Gärdenfors (eds.), *Applications of Conceptual Spaces*,
Synthese Library 359, DOI 10.1007/978-3-319-15021-5_4

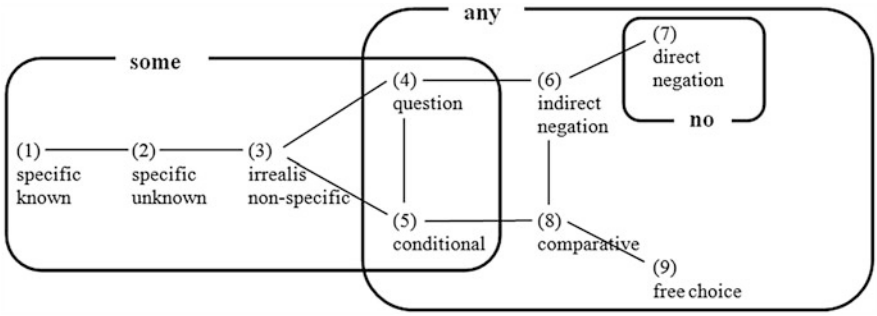

Fig. 4.1 Graph of functions of indefinite pronouns

The functions that indefinite pronouns can have are organized in a graph, with more similar functions closer to each other. Indefinite pronouns with *some* (e.g. *something*), *any* (e.g. *anything*), and *no* (e.g. *nothing*) correspond to contiguous sets of functions on the graph.

In one form or another this 'spatial' approach has been used for such diverse domains as modal verbs like *may* and *can* (Van der Auwera and Plungian 1998), container nouns like *jar* and *bottle* (Malt et al. 1999), motion verbs like *climb* and *crawl* (Geuder and Weisgerber 2002), adpositions like *in* and *on* (Levinson and Meira 2003), verbs of cutting and breaking (Majid et al. 2008), and case markers like the dative and accusative (Grimm 2011), to name just a few.

A conceptual space consists of a set of 'meanings' (like colors, referential functions, modalities, pictures of containers, pictures of spatial relations, video clips of cutting and breaking events, bundles of semantic properties) and some mathematical structure defined over that set.[1] This structure can be a discrete graph (like Fig. 4.1) or it can be a continuous metric (like the color space), but the idea is always that meanings that are closer together in the conceptual space are more similar, like the specific or negative functions in Fig. 4.1.

There are at least three ways in which one can construct such a similarity space for a domain:

- The lexical way: Meanings are closer if speakers use the same lexical item for them more often across languages.
- The psychological way: Meanings are closer if human subjects judge them to be more similar, non-linguistically.
- The semantic way: Meanings are closer together if a semantic analysis treats them as more similar.

The lexical way is the one most traveled. We find it in the graph-based semantic maps of typological linguistics (Haspelmath 2003) and in the statistical approaches

[1]Note that the term 'meaning' is used here in a very broad sense, encompassing both specific referents and general functions, and both non-theoretical and theoretical notions of meaning.

of Levinson and Meira (2003) and Majid et al. (2008). For example, in the space of Fig. 4.1, the meanings 'conditional' and 'comparative' are close together because languages tend to use the same word for those meanings, like English does with *any*. The psychological way of deriving a conceptual space can be seen in Malt et al. (1999) and in much other work that uses pile sorting or other ways to derive non-linguistic similarity judgments of stimuli. Van der Auwera and Plungian (1998), Geuder and Weisgerber (2002), Grimm (2011), and Gärdenfors et al. (2012) build their spaces on semantic considerations, that is, on an analysis of the functions or referents that are covered by the expressions that they study.

In Zwarts (2010), I argued that conceptual spaces need to be approached from different, complementary angles, especially if we want to study constraints on lexical categorization. If we want to understand how words can be meaningfully used and learned, then we need to understand what can and cannot be a word meaning. Gärdenfors (2000) and Haspelmath (2003) independently argued for *geometric* constraints on meaning. According to Gärdenfors, meanings are *convex*, while in Haspelmath's semantic maps they are *contiguous* (connected, in graph theoretical terms). Such constraints also play an important role in the modal map of Van der Auwera and Plungian and the case map of Grimm. Obviously, if we want to test whether the regions corresponding to words satisfy certain constraints, then those constraints should not themselves be part of the recipe for making the space, as in the lexical approach, but the space should be constructed independently of the lexical items that are distributed over it. Only then can we test the hypothesis that word meanings are convex or contiguous.

In a sense, this paper is an experiment. It starts with a lexical domain described in Geeraerts et al. (1994) (henceforth GGB) for which three things were given: (i) a set of referents (244 types of shirts), (ii) a set of words from one language applying to those referents (seven Dutch words), (iii) an analysis of each of these referents in terms of properties (shape, fabric, fastening, ...). The set of referents is too big in relation to the set of words to follow the lexical route to a conceptual space. There are only seven words and 244 referents, so we cannot induce the similarity relations of the referents on the basis of those words. But suppose we would use the feature analysis of the referents that GGB provide, what kind of space would we get, with what sort of 'shape', and how is the distribution of the words constrained by the geometry of the space? Answering these questions is what I set out to do in this paper. Although the results are not entirely conclusive (because of the nature of the data, as we will see), the approach looks promising because of the way it explicitly links constraints of categorization to semantic features, making them testable, in principle.

This paper is structured as follows. After introducing the domain of shirts in Sect. 4.2, I will define the 'shirt space' in Sect. 4.3. This space has a particular 'shape' that shows two major clusters of shirt types (Sect. 4.4). I then discuss the status of four different categorial constraints in this feature-based shirt space: conjunctivity (Sect. 4.5.1), convexity (Sect. 4.5.2), connectivity (Sect. 4.5.3), and coherence (Sect. 4.5.4).

4.2 Dutch Shirts

GGB report the results of research done at the University of Leuven, Belgium, between 1990 and 1993, into the nature and origins of lexical variation. A total of 9,000 occurrences of clothing terms were collected from magazines in Dutch, published in Belgium and the Netherlands. In each case the occurrence of the term was accompanied by a picture that showed an instance of the item. These pictures were used to make componential analyses of the referents of clothing terms.

For example, one of the referents encountered has feature decomposition [52131] and is found labeled by words like *spijkerbroek* and *jeans*, both of which mean 'jeans'. Each of the five digits of [52131] represents a value on a particular dimension:

Length: 5 (down to the ankles)
Width and cut: 2 (straight cut, neither tight nor wide)
End of legs: 1 (no special features)
Material: 3 (denim)
Details: 1 (strengthened by metal buttons)

There might be other referents, with different features that are also labeled *spijkerbroek* or *jeans*. This means that every clothing term has an extension that consists of referents, each uniquely described by a set of discrete features. It is important to realize that [52131] is not the traditional semantic decomposition of a *word meaning*, but the analysis of one particular *referent*.[2] GGB were interested in determining the prototype structure or family resemblance structure of clothing categories, not necessarily their classical definition in terms of necessary and sufficient features, which might not always exist.

There are not many details about how the researchers went about to make their feature decomposition and the original pictures are not included in the book. There are probably different ways in which one can analyze a set of clothing items into features, and certain decisions are made that would need further motivation, but nevertheless, I am assuming for now that their analysis into features is at least a plausible way to represent what the referents are like, coming back to possible shortcomings along the way and in the conclusion.

I focus on one particular subdomain from their book, which involves 'shirts, t-shirts, blouses; garments covering the upper part of the body, made of light material, constituting the first layer of clothing above the underwear' (p. 22). All the data are taken from their book (p. 129–133). They present a total of 244 configurations (referent types), that were analyzed for features of shape, length, fastening, fabric,

[2] A referent like [52131] is a *type* of object that corresponds to many different *tokens* that share these five properties, but differ in a lot of other properties.

collar, neckline, position of the buttons, and gender.[3] There were 7 Dutch terms found for these items, which are all covered more or less by the English noun *shirt*, each of which has as its extension a proper subset of the 244 shirt types. The following glossing is only a very rough approximation:

blouse 'blouse', *hemd* 'shirt', *overhemd* 'dress shirt', *overhemdblouse* 'shirtwaist', *shirt* 'shirt', *t-shirt* 't-shirt', *topje* 'tank top'

On the basis of the data they collected for these terms, GGB demonstrated that this field has a 'non-classical, un-mosaic-like character' (p. 134), without 'sharp divisions between the individual items within the field' (p. 118). The terms are overlapping to a large extent and they are not hierarchically ordered: *overhemd*, for instance, is not a hyponym of *hemd*, and *t-shirt* not of *shirt*.

The features with their values are as follows, each given with an example:

Shape: 1,2,3,4 (e.g. 4 = covering trunk and arms)
Length: 1,2,3,4 (e.g. 2 = tucked into skirt or trousers)
Fastening: 1,2,3,4 (e.g. 3 = full fastening)
Fabric: 1,2,3,4 (e.g. 1 = smooth, cottonlike)
Collar: 1,2,3,4,5 (e.g. 3 = soft collar)
Neckline: 1,2,3,4,5,6 (e.g. 3 = round neckline)
Position of the buttons: 1,2,3 (e.g. 2 = left)
Sex: v,m (e.g. v = female)

GGB have used discrete feature values for dimensions that are truly discrete, like the position of buttons (left, right) or the gender of the shirt (male, female), but also for dimensions that are really continuous, like length, by partitioning this continuous scale into a small number of intervals.

Each particular shirt corresponds to a string of feature values. Referent number 3, for instance, corresponds with the string [4231332v]. Following a common linguistic practice, I put square brackets around a feature bundle. I refer to referents with S1, S2, etcetera, in order not to confuse them with feature values. The ordering of the values of a feature is not significant, although in some cases the choice of integers is not entirely arbitrary. For instance, the other values of Shape are 1 = covering trunk below shoulders, 2 = covering trunk and shoulders, leaving arms uncovered, and 3 = covering trunk, shoulders and upper arms. I will treat the values of all features as unordered.

Because we are working with strings, it is possible to pick out classes of shirts with regular expressions in the usual way, as a useful notion. The full stop (.) is used as the wildcard for any value of a feature, the vertical bar (|) for alternative values, and the dash (-) for a range of values. The strings of features can now be used to define our 'shirt space', which is the topic of the next section.

[3]The book actually gives 246 configurations, but there are two pairs with the same feature profile, so I counted both of these pairs as one referent. Therefore the configurations numbered 67 and 68 are missing in my list.

4.3 Shirt Space

The idea of a conceptual space is to represent semantic similarity of the elements of a domain in terms of spatial distance. For our situation this means that we have to determine distances between shirts on the basis of their feature makeup. For this I use a very simple metric, namely the *Hamming distance*.[4] If we consider only strings of equal length, then the Hamming distance between two strings is simply the number of positions at which they differ. For instance, the distance between configuration S1, characterized by the string of features [3231432v], and configuration S2, with the string of features [3431412v] is 2, because they differ only in the second and sixth position. For short, I will use $d(x,y)$ for the distance between two feature bundles x and y in a set of such bundles S. Given the way distance is defined, we have the following properties for all configurations x, y, and z in S:

$$d(x, y) \geq 0$$
$$d(x, y) = 0 \text{ iff } x = y$$
$$d(x, y) = d(y, x)$$
$$d(x, y) + d(y, z) \geq d(x, z)$$

This makes our set of shirts a metric space, in which similar shirts are closer together (see Gärdenfors 2000 for the metric properties of conceptual spaces).

Note that this similarity metric is very simple. It does not take into account that some features (like Sex) have two values, while other features (like Neckline) have six. The distance between a female and male version of a shirt is just as great as the distance between a shirt with a round or rectangular neckline. In this similarity metric all the features have the same weight, to keep things simple, but also because it is not straightforward to determine what the weights would have to be for this set of data.

In a sense, this approach follows the opposite direction from approaches that start with lexical or similarity judgment data and perform multidimensional scaling or a similar statistical operation to derive the features. By starting with features, instead, the dimensions of the space are already given, because each feature with its range of values can be seen as representing a dimension. We usually think of a dimension as a continuous scale, but here the dimensions only have a few discrete and unordered values. It is on the basis of these 'a priori' given features/dimensions that a metric space is defined over the set of referents. Multidimensional scaling works in the opposite direction. It starts with a high-dimensional metric space for a set of referents (based on how often referents are named by the same word across languages, for instance), and then extracts a few dimensions that best represent this high-dimensional space.

[4]For more sophisticated similarity measures based on features see Tversky (1977).

This distance metric d can be used to define a notion of *betweenness b*.

For all distinct x, y, z in S, y is between x and z if and only if $d(x,y) + d(y,z) = d(x,z)$.

So, for instance, referent S107 [3431432v] is between S1 [3231432v] and S2 [3431412v], because $d(\text{S1,S107}) = 1$ and $d(\text{S107,S2}) = 1$ and $d(\text{S1,S2}) = 2$. We can say that two distinct elements are *adjacent* when there is no element between them:

For all distinct x and z in S, x is adjacent to z if and only if there is no y in S such that y is between x and z.

Referent S1 and S2 are not adjacent to each other, but they are both adjacent to S107. Adjacent referents can have a distance greater than 1: S1 and S102 [3231431m] have a distance of 2, but they are adjacent because there are no referents with values [3231432m] or [3231431v].

Finally, we can define the notion of a *path* in a space S:

For all distinct x and y in S, a path from x to y is a sequence of distinct elements $x,y_1,\ldots,y_n,z$ in S with $n \geq 0$, such that every two subsequent elements are adjacent.

For example, one path from S1 [3231432v] to S4 [4231411v] is the sequence S1,S6,S109,S4, where S6 is [4231432v] and S109 is [4231412v]. In this case, each step of the path corresponds with one feature difference. There can of course be more paths between two referents and the distances between the referents on the path can be two or more.

With this feature decomposition, we can now try to address the questions that we posed earlier. What kind of constraints can we find on lexical regions defined over a feature-based space? But first we look at the shape of the space itself that is defined by the features.

4.4 The Shape of the Space

With the number of features and values presented in the preceding section, a total of 38,400 possible items can be defined. However, there are only 245 actual distinct items in this data set, which means that we are dealing with an irregularly shaped semantic space, in which some areas are more populated than other areas. This might have partially to do with the inevitable restrictions of sampling: certain types of shirts might exist but were simply not found in the magazines used for the sample. Also, the sample was biased to women's clothing, because of the nature of the magazines. But there are definitely also some real constraints on the space.

- Cultural constraints, having to do with the kind of shirts that are worn by men and women. For example, those shirts that only cover the trunk below the shoulders are exclusively worn by women (in other words, there are no referents [1.....m] in the database, where 1 means 'only covering the trunk below the shoulders' and m is 'worn by men').

- Constraints of a more logical nature. If a shirt is specified as having no fastening, then the buttons are neither right nor left (so there are no referents with the feature description [..1...(1|2).]).
- Physical (or 'technical') constraints. A wide neckline, for instance, does typically not allow for a collar. (Of the twelve [.....6..] shirts, eleven are [....16..] and one is [....36..]). 6 is the wide neckline feature, 1 is the no collar feature.

In this way, one might expect the shirt space to have an irregular shape, in some general sense like what has been argued for the color space (Regier et al. 2009), which has different saturations for different hue-lightness combinations. However, unlike the color space, the shirt space does not owe its irregularity to perceptual factors, but to other factors. Unfortunately, there is not enough information in GGB to pursue this topic more. Therefore, we have to look at more general ways to look at the shape of the space.

There are different ways in which one can spatialize and visualize a space like this. The graph in Fig. 4.2 shows all the shirts, but draws only edges between two shirts if they differ in exactly one feature.[5] Most of the shirts are connected in this way, but some are floating around unconnected, simply because their distance to other shirts is greater than 1. They are randomly placed by the drawing program.

Notice that we do not get an explicit representation of the dimensions of the space in this way. These dimensions are spread out, in a sense, across the graph, as we can see when we give the vertices of the graph a color that represents with a particular feature value. For instance, in Fig. 4.3 the four different values of the feature Shape are distinguished by color in the following way (descriptions taken from GGB, p. 129):

Blue: Covers the trunk below the shoulders ([1.....]).
Red: Covers trunk and shoulders but leaves the arms uncovered ([2.....]).
Green: Covers trunk, shoulders and upper arms, but leaves the lower arms uncovered ([3.....]).
Yellow: Covers the trunk and the arms ([4.....]).

We can only make the feature dimensions more explicit if we take 'cross sections' of the shirt space. A two-dimensional cross section might consist of the features Sex (with values m, v) and Shape (with values 1, 2, 3, 4), as shown in Fig. 4.4. The values of Shape are ordered from covering less to more of the upper body, which happens to correspond to the ordering of the corresponding integers. The combination [1.....m], here abbreviated as m1, as we already saw (there are no shirts for men that only cover only the trunk below the shoulders).

Two clusters present themselves in the space of shirts in Figs. 4.2 and 4.3. When we inspect how the features of those two clusters differ, then we actually see that those clusters correspond roughly to two types of shirts. In Fig. 4.5 two broader types of shirts are indicated. Green corresponds to the more formal type

[5]The graph is drawn by the graphviz software package, using the neato command, which creates a layout that approximates distances in the graph.

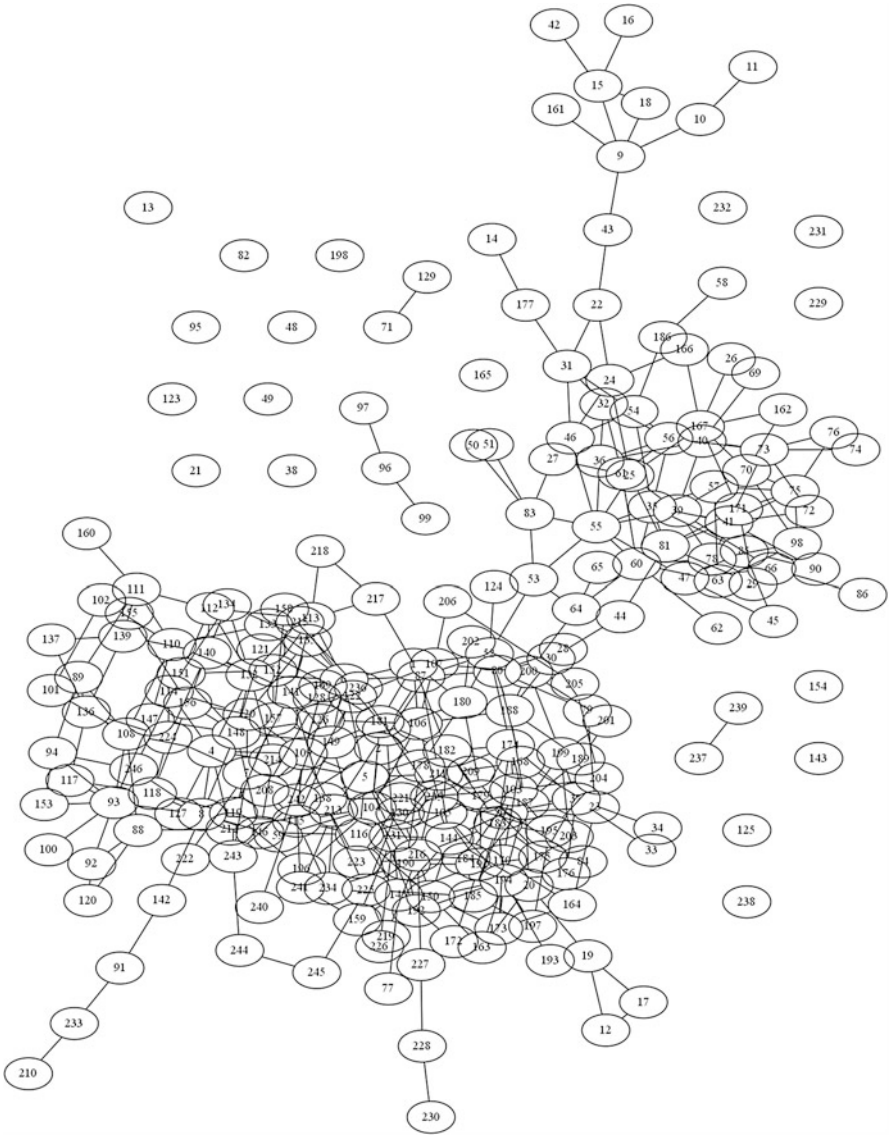

Fig. 4.2 Shirts with one feature difference linked

of dress shirts, with a full button fastening and a collar ([..3.[2-5]...]), while red corresponds to the less formal type of shirt, lacking buttons and collar ([....1.3.]).[6]

[6]Notice that graphviz does not draw the same graph with a fixed orientation, which is why the graph in Fig. 4.5 is rotated.

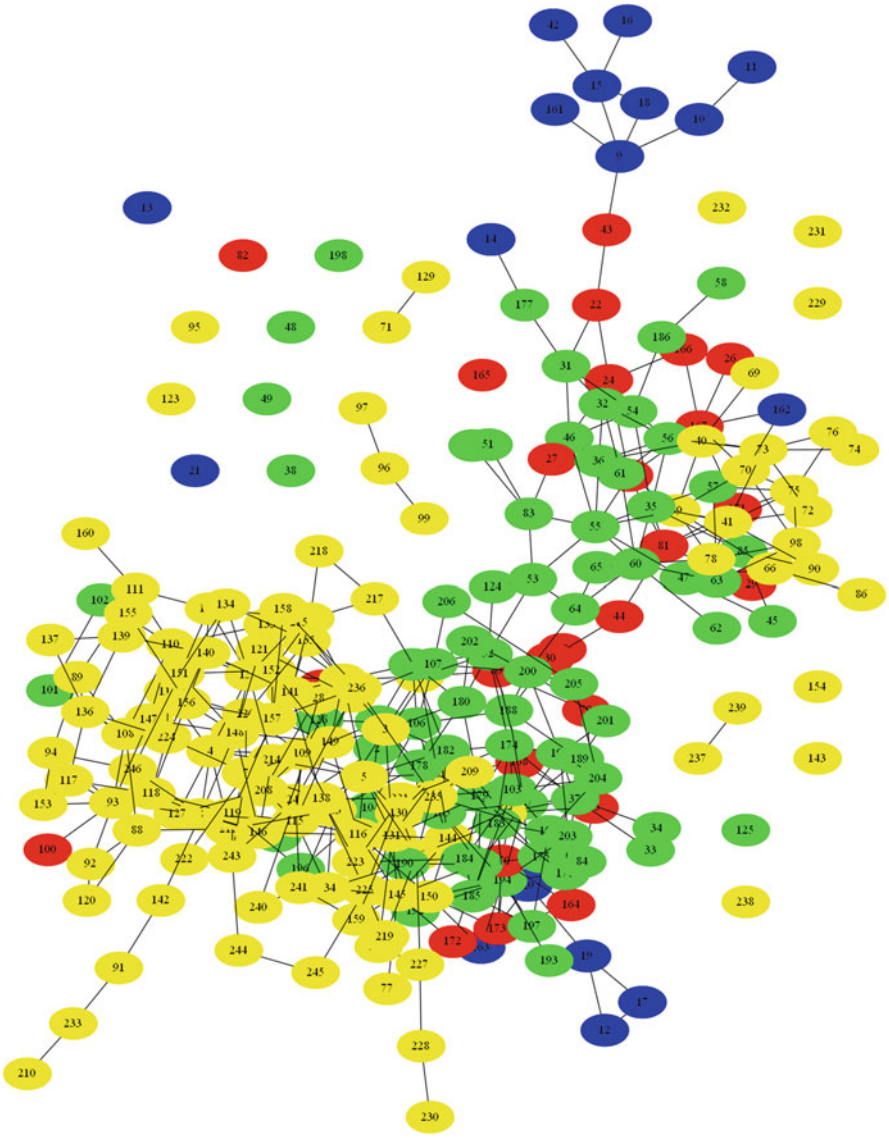

Fig. 4.3 Distribution of the shape feature

The two clusters seem to have a kind of feature-based identities. To what extent is that reflected in the naming patterns? Unfortunately, we can not represent all the seven shirt nouns together in one graph, because they overlap each other and hence do not partition the space. Figure 4.6 shows how three of the nouns distribute their extension over the space. Yellow is *topje*, red is *t-shirt*, blue is *overhemd*, orange is used for the shirts that are labeled both *topje* and *t-shirt* and purple is used for the shirt that is both named *t-shirt* and *overhemd*.

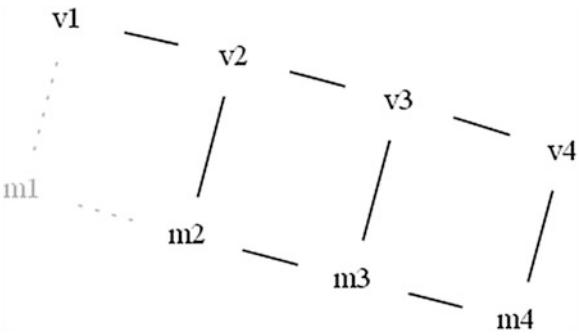

Fig. 4.4 Two-dimensional cross section of the shirt space

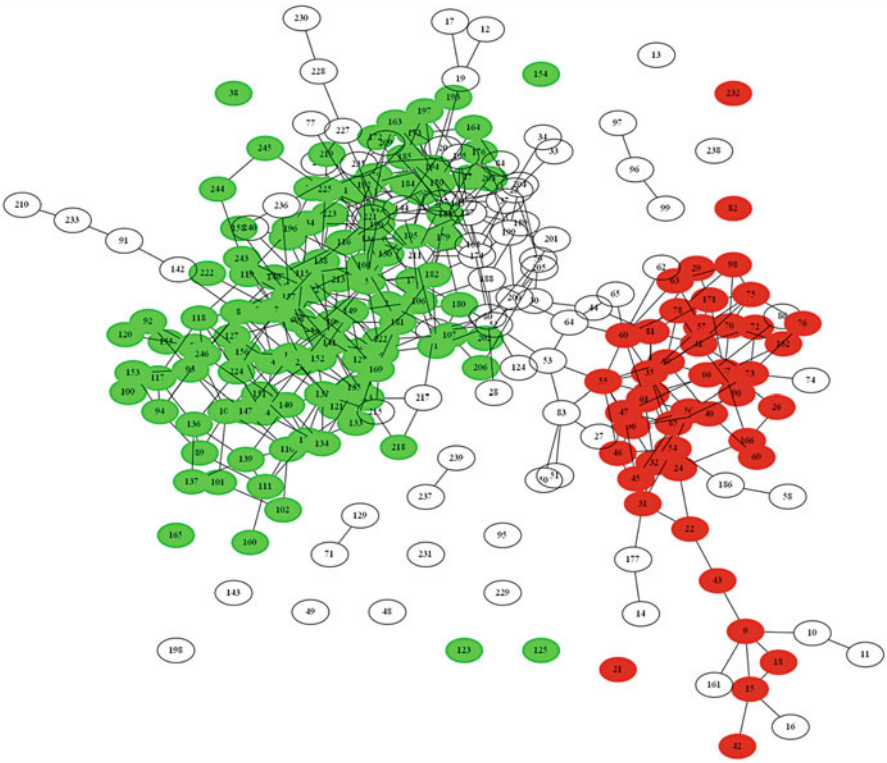

Fig. 4.5 Two types of shirts and their correspondence to clusters

As we can see, the names correspond quite well to the clusters. The items called *overhemd* (coloured blue) cluster in the more formal area of the space, while the items called *t-shirt* (red) and *topje* (yellow) cluster in the less formal area. Some members of *topje* are found in the *overhemd* cluster, which on closer inspection turn out to be ones that have a full fastening with buttons.

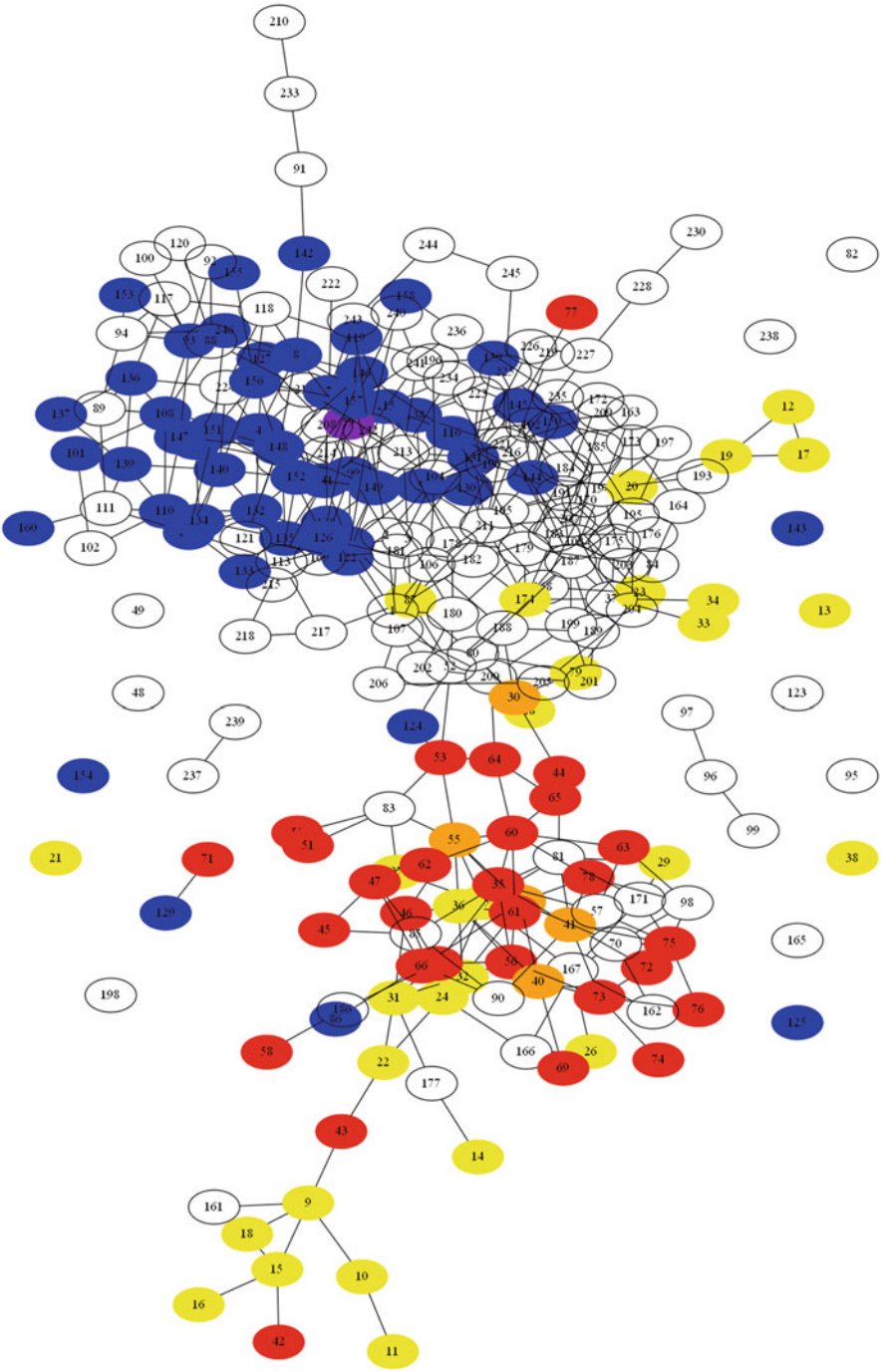

Fig. 4.6 *Overhemd*, *topje*, and *t-shirt* in the space of shirts

After this exploration of the structure of the space, we now turn to different ways in which the shirt categories might be constrained in terms of the underlying, feature-based space.

4.5 Constraints on the Categories

4.5.1 Conjunctivity

One theoretical possibility is that the shirt categories are classically definable by means of necessary and sufficient conditions, that is, as a conjunction of one or more features. We have already seen such categories in the previous section. For example, when we are talking about shirts with a collar and a full fastening with buttons or shirts that do not have a fastening, buttons or a collar, then we are using conjunctive definitions.

With a collar and a full fastening with buttons: [..3.[2-5].[1-2].]
Without a fastening, collar or buttons: [..1.1.3.]

For every feature, there is either no specification (.) or a range of one or more values (like [2-5]). A disjunctive definition would be a class of shirts that are either collarless *or* that have buttons, i.e. [....1...] $\vee$ [.....[1-2].], with two regular expressions.

As we know, conjunctive definitions make a lot of intuitive sense and they play an important role in certain domains (see Hage 1997). At the same time, one of the reasons that GGB undertook their empirical study of clothing terms is that they wanted to demonstrate that not all clothing terms allow for a conjunctive definition, but that prototype and family resemblance structure play an important role. In fact, it is not possible for the set of data that we have from GGB to come up with conjunctive characterizations of any of the shirt nouns. In order to cast the net wide enough to cover all the positive referents of a noun, we get also referents in our net that are not attested with that noun. For example, when we look at items S6 ([4231432v]) and S133 ([4231432m]) in the extension of *overhemd* we see that the sex for the person for whom the shirt is meant should not matter. So for every other female referent of *overhemd*, the corresponding male version should also be in the extension (if it exists). In other words, *overhemd* should be unspecified for sex. However, we find pairs of referents differing only in this feature and which are not both in the extension of *overhemd*:

S136: [4234411m] (*overhemd*) versus S224: [4234411v] (*blouse*)
S111: [4232431m] (*hemd*) versus S112: [4232431v] (*overhemd*)

However, I hesitate to draw definite conclusions from this about whether shirt categories might be conjunctively defined. First of all, we have to realize that, however rich the data are, they might still not be exhaustive. It could very well

be that with a bigger corpus of magazines, *overhemd* would have been attested for referent S224 and for referent S111. Asking mother tongue speakers of Dutch how acceptable they find *overhemd* for the relevant pictures might also have given different results. This point is also relevant for other constraints that we discuss in the next three sections. Second, it could be that *overhemd* has a classical, conjunctive definition, but that part of its extension is blocked by other words (*blouse* or *hemd*) that are more appropriate for that part. This is also something that can only be ascertained by a very large corpus or by deliberate elicitation.

Nevertheless, as far as the data go, the conclusion must be that conjunctivity is not a constraint that holds of word meanings in this domain.

4.5.2 Convexity

The constraint of convexity has been proposed by Gärdenfors (2000) as a constraint on natural properties, conceived as regions in a particular integrated domain. Even though it is not clear whether shirt names should be seen as referring to natural properties, still I believe it is worthwhile to investigate whether the extensions of the nouns are convex given the underlying feature space. The definition of convexity is as follows:

A subset C of a space S is convex if and only if for every x and y in C, all points between x and y are also in C.

Intuitively, convexity would make a lot of sense for shirt categories. If we take the shirts number S30 = [2331132v] and S35 = [3311133v], which are both called *t-shirt*, then what they have in common can be written as [.3.113.v]. The idea of convexity is that every referent that has these specific feature values and shares its other features with either S30 or S35, should also be called *t-shirt*. The referents that we find between S30 and S35 are S44 = [2331133v], S81 = [2311133v]), and S188 = [3331132v]. Of these three, only S44 is called *t-shirt*, while S81 is called *shirt*, and S188 *blouse*. So, already in this randomly picked example, convexity does not seem to work, but we have to look at it in a more general way.

We can get an idea of the extent to which categories are convex in this domain by using the notion of a *convex hull*, the closure of a set under betweenness.

In a space S, the convex hull H of set E is the union of E with those elements of S that are between members of E.

Here is a small example. Suppose that E = {S1 [3231432v], S2 [3431412v]}. There are two referents between S1 and S2, namely S107 [3431432v] and S181 [3231412v]. As a result, H = {S1, S107, S181, S2}. In a sense, we make the convex hull by filling up the 'hole' between S1 and S2.

It turns out that the convex hull is quite a bit bigger than the extension, for all of the nouns. In other words, there are quite a lot of 'holes' in the extensions. The numbers are as shown in Table 4.1.

Table 4.1 Non-convex complements of categories

Noun	Extension	Convex hull
blouse	116	210
hemd	30	127
overhemd	56	132
overhemblouse	7	37
shirt	25	141
t-shirt	37	120
topje	35	100

Every item that is in the convex hull of a category, but not in its extension is a counterexample against convexity. This clearly shows that these categories are not convex in the sense defined here. Again, different interpretations of this result are possible. As we already saw in the previous section, the corpora might not have provided enough naming data, thereby creating these 'holes'. Another response could be that it is not the category as a whole that is convex, but that it is only convex in certain dimensions. For example, clothing categories might be convex on the dimensions of shape. This seems also more in line with the position in Gärdenfors (2000). Yet another response might be that convexity is too strong a constraint here. What we need instead is *star convexity*. The category is then organized around a central point (a prototype) such that every referent in the category can be connected with this prototype by a line that is entirely within the category. This idea would be in line with a historical development of categories, from a prototype in different directions of similarity, maybe like the *chaining* in Malt et al. (1999). I leave it to further research to investigate this possibility.

4.5.3 Connectivity

While convexity is the constraint in Gärdenfors (2000), what we find in the semantic maps of Haspelmath (2003) is a weaker property of contiguity or, rather, connectivity in graph theoretical terms. The idea is that a word forms a connected subgraph of an underlying conceptual space. Maybe the shirt nouns are connected in this sense on the feature-based graph?

In order to investigate this, we need to construct a graph with the right kind of connections, for which adjacency seems appropriate. Remember that two meanings are adjacent if and only if there are no other meanings between them, which comes close to the kind of relation that underlies connectivity in semantic maps. With this relation, most of the shirt categories seem completely connected, like *overhemd* in Fig. 4.7, for instance, but unlike *hemd* in Fig. 4.8, which has two members (colored red) which are not adjacent to any other member. These diagrams show only the members of the categories together with their adjacency structure.

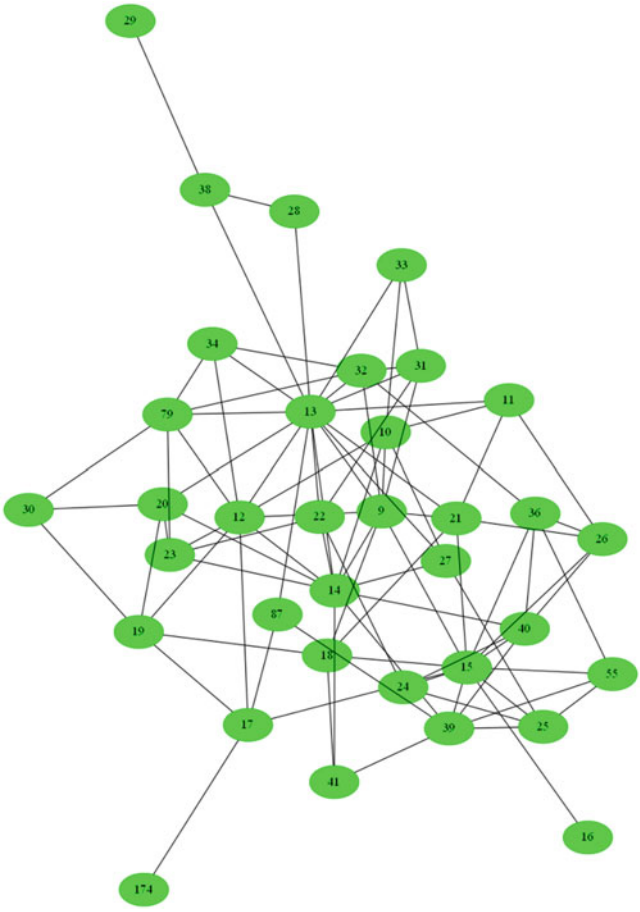

Fig. 4.7 The connectivity of *overhemd*

The problem is however that the kind of connectivity that adjacency gives us is too liberal. The reason is that every referent has quite a lot of adjacent neighbours. The number ranges from 5 to 63, with an average of a little bit over 17. This means that even if we would form an arbitrary set of elements, then the chance of each member to be adjacent to at least one other member of that set is quite big. The reason that our shirt space behaves this way might be that the theoretical space of possible types is quite large, as we saw, but because there are many gaps in the space, there are adjacencies over long distances, making this a quite tightly knit space, as a whole. An alternative way of defining adjacency would be in terms of Hamming neighbors, referents that differ only in exactly one feature. However, because of the sparsely filled space, this version will tend to underestimate connectivity.

Let me turn to the last possibility.

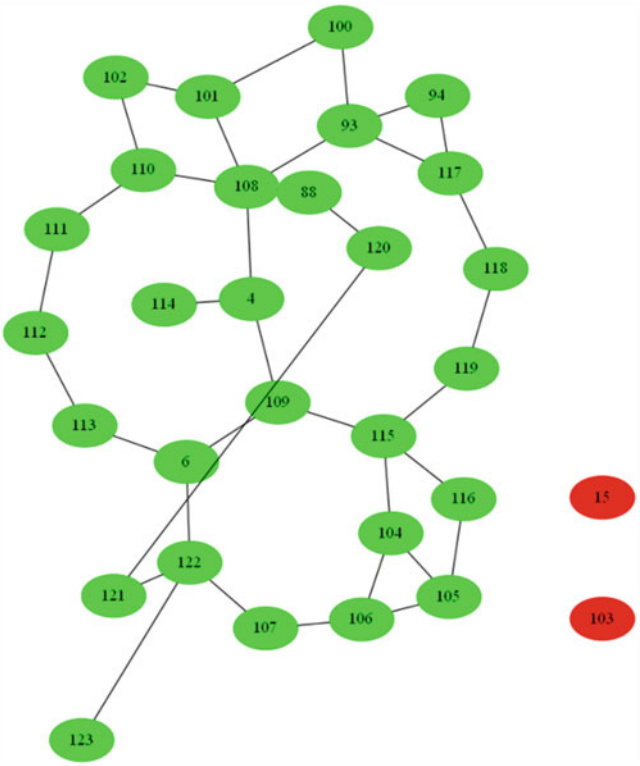

Fig. 4.8 The connectivity of *hemd*

4.5.4 Coherence

A common idea in the literature is that categories are 'coherent' or 'compact', meaning that they maximize the within-category similarity and minimize the across-category similarity (Rosch and Mervis 1975; Tversky 1977; Regier et al. 2009, among others). These last authors even showed that the partitions of the colour space by languages are near-optimal in the sense that for a whole partition the well-formedness is higher than for alternatives that are derived by rotating the color space with respect to a particular set of terms. Such rotations yield alternative terminological systems that are closely related to the original. A full exploration of this idea for the shirt space would go too far for this paper, because it is not immediately clear how one would go about defining rotations of the shirt space to derive close related alternative categorizations.

Nevertheless, 'rotation' in a looser sense is an easy way to shift the extension of a category to a more random alternative category. Suppose we shift a given category to a new category by shifting all its referents up in the list by a particular number,

Table 4.2 Average distance within categories and their shifted versions

Noun	No shift	+10	+20	+30	+40	+50
blouse	4	4.2	4.3	4.4	4.5	4.5
hemd	3.6	4.1	3.9	3.9	4	4.6
overhemd	3.4	4.5	4.4	4.2	4.1	4
overhemdblouse	2.5	3	2.8	3.4	3.4	2.8
shirt	4.7	4.8	4.2	4	3.9	3.5
t-shirt	3.5	3.8	4.3	4.7	4.7	4.3
topje	3.7	4.1	3.8	3.7	3.9	4.4

going by the numbering of GGB.[7] We can compute the average distance among the members of a set of shirts in the following way:

If S is a set of n items with a distance metric d, then the average distance $D(S)$ of S is given by $\Sigma\, d(x,y)$ for each pair $x,y \in S$, divided by $½n(n-1)$.

What we can see in Table 4.2 is that the average distance of categories that are shifted (over five different distances) is generally higher than the average distance of the original category, which supports the idea that categories have some sort of coherence. However, there is one interesting counterexample, namely *shirt*. The average distance within this category is a bit higher than within the other categories and some of the shifts of *shirt* make the category actually more coherent it seems. One interpretation might be that *shirt* does not have a well-established shared meaning across different users and therefore lacks in coherence.

Let me finish this section by showing how we can get a visual impression of coherence when we display the categories in a graph that also respects the distances between the nodes. In this display, multidimensional scaling is used as a model for approximating the distances between the nodes. As we can see in Fig. 4.9, the region corresponding to *hemd* is fairly coherent, with only one clear 'outlier'. The red nodes are the disconnected ones, in the sense described earlier. Figure 4.10 shows that *shirt* distributes over the space in a much less coherent way.

It is also possible to identify members of a category that minimize the average distance to other members, and are central in that sense. For the category *hemd*, referent S4 has the smallest average distance to other members. One might want to say that S4 is like the prototype of this category. It should have features then that are more typical of this category and this actually turns out to be true. S4 has the profile [4231411v] and when we look at the feature values that occur most frequently with referents of *hemd*, as shown in Table 4.3, then we can see that most frequent values for each feature (highlighted by boldface) are exactly the feature values of the referent that is spatially central, thereby reflecting the redundancy structure of the category as a whole (Rosch and Mervis 1975).

[7]There is no system in the list, apart from the fact that sometimes a contiguous range of items in the list seem to belong to the same category.

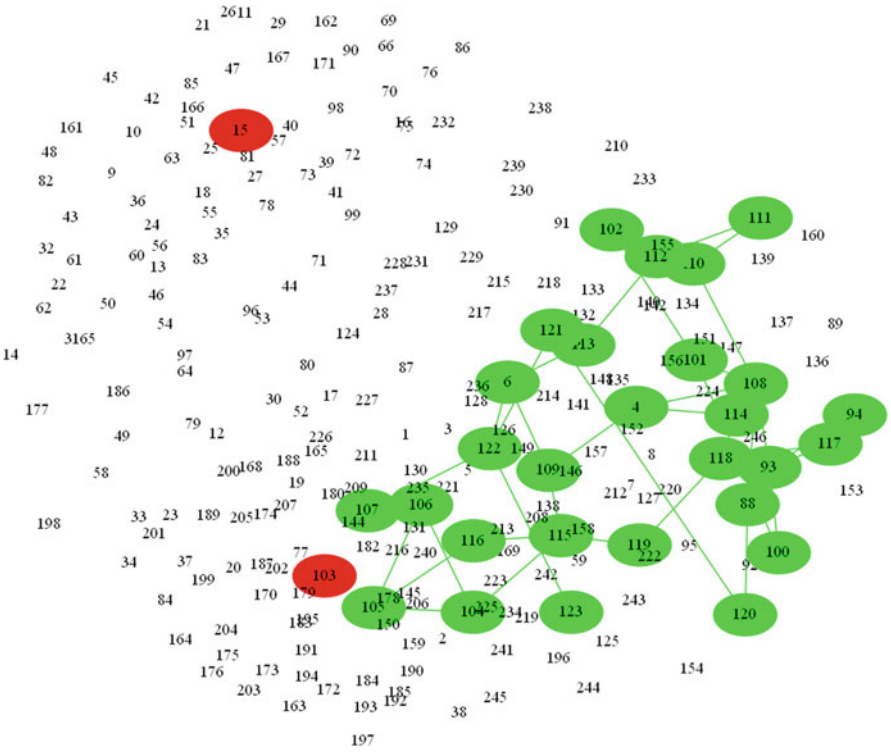

Fig. 4.9 The coherence of *hemd*

4.6 Conclusion

This paper has explored how we can construct a conceptual space on the basis of a given set of features and thereby study different properties of linguistic categories. For this particular domain of shirts and Dutch data set, the patterns might be linked to what has been found in another domain of artefacts, namely (household) containers (e.g. Malt et al. 1999, 2010), where the link between the perceived features and the linguistic labelings seems fairly 'loose' and strongly influenced by language-specific and culture-specific factors.

We have also seen that the extensions of shirt nouns show coherence in terms of the underlying feature space, but that they do not show the kind of conjunctivity or convexity that we would expect if they were based more directly on concepts (like 'long-sleeved shirt with a stiff collar with full fastening with buttons'). It is conceivable that these categories are not only held together on the basis of the underlying space, but also by conventions that cause referents to belong to the same category, even if they are not spatially related in one way or another. We know that metaphorical and metonymical mappings can extend the application of a term in a 'non-local' way and similar mechanisms might be at work in this domain

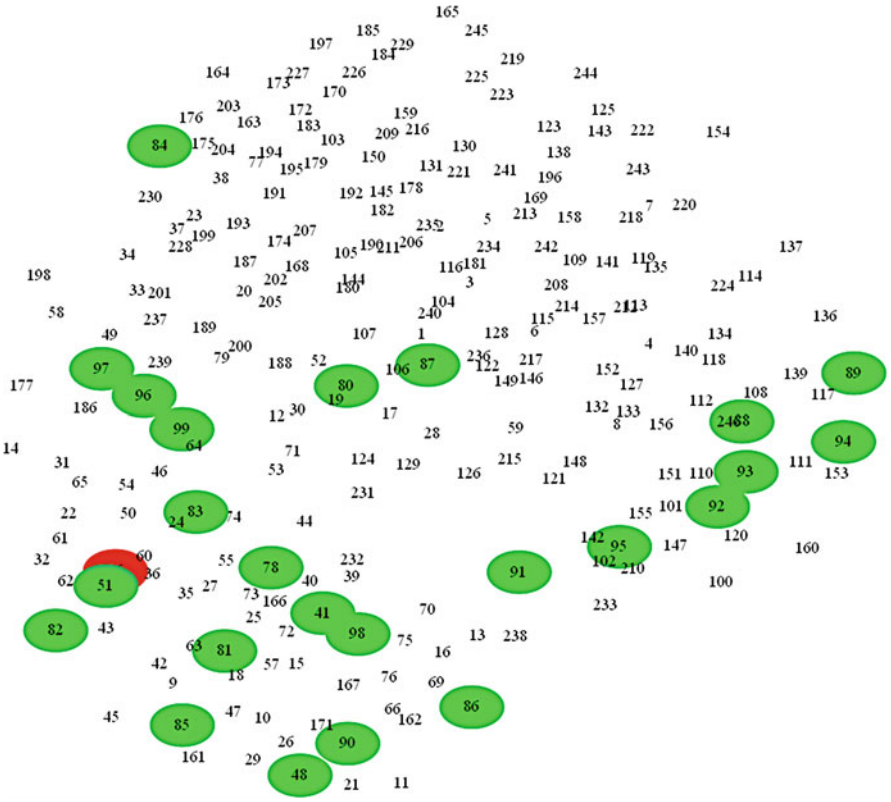

Fig. 4.10 The lack of coherence of *shirt*

Table 4.3 Number of referents per feature value for *hemd*

Feature	1/v	2/m	3	4	5	6
Shape	1	1	7	**21**	–	–
Length	0	**14**	11	5	–	–
Fastening	1	0	**29**	0	–	–
Fabric	**20**	4	1	5	–	–
Collar	2	0	3	**25**	0	–
Neckline	**16**	3	11	0	0	0
Buttons	**16**	13	1	–	–	–
Sex	**19**	11	–	–	–	–

too. However, there are several considerations that need to be kept in mind when evaluating the categorial constraints, having to do with the nature of the data and with the nature of the constraints.

The data of this study come from a corpus of fashion magazines. Although GGB made sure that the corpus was saturated in the sense that a bigger corpus would not have contained more *word types* or *referent types*, still a bigger corpus might have

given more *naming relations*, that is, applications of words to referents. If a referent *r* is not named by noun *n* in the corpus, then this does not allow us to conclude that certain strong constraints, like convexity, do not hold. Elicitation of naming relations directly from native speakers seems a better way of collecting the relevant data, but even here care should be taken that acceptability of a noun *n* for a referent *r* is tested exhaustively. Even in the experimental studies of naming there is a tendency to go for the most frequent label of a referent, which is not a good measure if we want to study strong constraints on lexical categories.

Another problem of corpus data for studying general constraints on categorization is that they might be constituted of rather different language varieties, at the level of idiolects, sociolects, or dialects. The corpus of GGB was deliberately composed in such a way that such variation could be studied. However, the result could be that a noun lacks a particular property (like convexity) because its extension in the data set is the union of two or more different uses of that noun, each with their own extension. While coherence might still hold of such an aggregated extension, properties like conjunctivity or convexity are better treated as constraint on the categorizations of individual language users.

As we saw, the elements of the domain are all very close together, leading to an overall high level of connectedness. It seems more likely that features or feature values do not all contribute to the structure of the space and the constitution of categories in the same way. Certain features are more salient than others and will play a greater role in categorization. As Tversky (1977) already showed, features can be weighed, and this might affect the results and conclusions in important ways. The question is then, of course, how one could define weights for the features that define a shirt.

This question is part of the more general question of how the features can best be defined and motivated. Which features are used by humans to categorize shirts and what determines the salience of those features in their perception and categorization? What kind of values can a feature take, ranging from binary (male/female user) to multidimensional and continuous (the shape)? How do those feature values affect the shape of the conceptual space and the distances between referents in that space? I have shown that a conceptual space and spatial constraints can be meaningfully defined on the basis of discrete features, but the results might be different if continuous dimensions are used where possible.

Finally, there are more sophisticated methods and techniques that could be used to study feature-based spaces, such as network analysis and machine learning, and all the statistical and graphical methods that are part of that. In that way, inductive and deductive, discrete and quantitative, theoretical and empirical approaches to word meaning can be more tightly integrated within the general perspective of conceptual space semantics.

Like I said at the beginning, this paper presents an experiment in the exploration of a conceptual space. Although it does not yield definite results about how concepts in conceptual space might be constrained in terms of the underlying features, it does show the usefulness of such an approach, especially if this linguistic approach would be wedded with both computational and psychological methods.

Acknowledgements I thank the audience for helpful questions and remarks as well as two anonymous reviewers and Peter Gärdenfors for their comments. The Netherlands Organization for Scientific Research (NWO, grant 360-70-340) and the Swedish Collegium for Advanced Study are also gratefully acknowledged for their support.

References

Gärdenfors, P. (2000). *Conceptual space: The geometry of thought*. Cambridge, MA: The MIT Press.

Gärdenfors, P., Warglien, M., & Westera, M. (2012). Event structure, conceptual spaces, and the semantics of verbs. *Theoretical Linguistics, 38*(3–4), 159–193.

Geeraerts, D., Grondelaers, S., & Bakema, P. (1994). *The structure of lexical variation: Meaning, naming, and context*. Berlin/New York: Mouton de Gruyter.

Geuder, W., & Weisgerber, M. (2002). Verbs in conceptual space. In G. Kath, S. Reinhard, & P. Reuter (Eds.), *Sinn & Bedeutung VI proceedings of the sixth annual meeting of the Gesellschaft für Semantik* (pp. 69–83). Osnabrück, Germany: University of Osnabrück.

Grimm, S. (2011). Semantics of case. *Morphology, 21*, 515–544.

Hage, P. (1997). Unthinkable categories and the fundamental laws of kinship. *American Ethnologist, 24*(3), 652–667.

Haspelmath, M. (1997). *Indefinite pronouns* (Oxford studies in typology and linguistic theory). Oxford: Oxford University Press.

Haspelmath, M. (2003). The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In M. Tomasello (Ed.), *The new psychology of language* (Vol. 2, pp. 211–242). Mahwah: Lawrence Erlbaum.

Kay, P., Berlin, B., Maffi, L., Merrifield, W. R., & Cook, R. (2009). *The world color survey*. Stanford: Center for the Study of Language and Information.

Levinson, S. C., & Meira, S. (2003). 'Natural concepts' in the spatial topological domain – Adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language, 79*(3), 485–516.

Majid, A., Boster, J. S., & Bowerman, M. (2008). The cross-linguistic categorization of everyday events: A study of cutting and breaking. *Cognition, 109*(2), 235–250.

Malt, B. C., Sloman, S. A., Gennari, S., Meiyi Shi, & Yuan Wang. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language, 40*, 230–262.

Malt, B. C., Gennari, S., & Imai, M. (2010). Lexicalization patterns and the world-to-words mapping. In B. C. Malt & P. Wolff (Eds.), *Words and the mind: How words encode human experience* (pp. 29–57). Oxford: Oxford University Press.

Regier, T., Kay, P., & Khetarpal, N. (2009). Color naming and the shape of color space. *Language, 85*, 884–892.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573–605.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327–352.

Van der Auwera, J., & Plungian, V. (1998). Modality's semantic map. *Linguistic Typology, 2*, 79–124.

Zwarts, J. (2010). Semantic map geometry: Two approaches. *Linguistic Discovery, 8*(1), 377–395.

Chapter 5
Meaning Negotiation

Massimo Warglien and Peter Gärdenfors

Abstract While "meaning negotiation" has become an ubiquitous term, its use is often confusing. A negotiation problem implies not only a convenience to agree, but also diverging interest on what to agree upon. It implies agreement but also the possibility of (voluntary) disagreement. In this chapter, we look at meaning negotiation as the process through which agents starting from different preferred conceptual representations of an object, an event or a more complex entity, converge to an agreement through some communication medium. We shortly sketch the outline of a geometric view of meaning negotiation, based on conceptual spaces. We show that such view can inherit important structural elements from game theoretic models of bargaining – in particular, in the case when the protagonists have overlapping negotiation regions, we emphasize a parallel to the Nash solution in cooperative game theory. When acceptable solution regions of the protagonists are disjoint, we present several types of processes: changes in the salience of dimensions, dimensional projections and metaphorical space transformations. None of the latter processes are motivated by normative or rationality considerations, but presented as argumentation tools that we believe are used in actual situations of conceptual disagreement.

5.1 Introduction

"Meaning negotiation" has become an ubiquitous term, used in contexts as diverse as semantics and epistemology (Larson and Ludlow 1993), conversation theory (Brennan and Clark 1996), ethnography (Wenger 1998), but also literary criticism, artificial intelligence, psychotherapy. The concept suggests that meaning is often not

M. Warglien (✉)
Center for Experimental Research in Management and Economics,
Ca Foscari University, Venezia, Italy
e-mail: warglien@unive.it

P. Gärdenfors
Department of Philosophy and Cognitive Science, Lund University, Box 192,
221 00 Lund, Sweden
e-mail: peter.gardenfors@lucs.lu.se

F. Zenker, P. Gärdenfors (eds.), *Applications of Conceptual Spaces*,
Synthese Library 359, DOI 10.1007/978-3-319-15021-5_5

uniquely determined by the lexicon and ordinary utterances, and thus there is room left for a process of further determination through some type of interaction among communicating agents.

However, the expression "meaning negotiation" may be sometimes a source of confusion. In many current usages of the term, a negotiation is often confused with an agreement. However, a negotiation problem implies not only a convenience to agree, but also diverging interest on what to agree upon. It implies agreement but also the possibility of (voluntary) disagreement. Thus, a problem of negotiation differs from a problem of pure coordination, since while negotiators both have an interest to agree (as in coordination), they nevertheless have conflictual interests in dividing the value generated by their cooperation. Schelling (1960) has described this type of interaction as one of "mixed motives", since common interest and conflict coexist in the same situation.

Robert Stalnaker has well captured the issue of "mixed motives" in his description of a conversation game:

> One may think of a non-defective conversation as a game where the common context set is the playing field and the moves are either attempts to reduce the size of the set in a certain ways or rejections of such moves by others. The participants have a common interest in reducing the size of the set, but their interests may diverge when it comes to the question of how it should be reduced. (Stalnaker 1999)

In this chapter, we will look at meaning negotiation as the process through which agents starting from different preferred conceptual representations of an object, an event or a more complex entity, converge to an agreement through some communication medium. The process is typically a sequence of offers and counter-offers that are accepted or rejected as in Clark's "contributions" (Clark and Schaefer 1989). The "solution" to the negotiation problem is the agreement reached (or the final disagreement). While this approach maintains a broad scope, it is important to stress that it assumes that agents "move" in a defined conceptual space and have potentially conflicting interests in the agreement to be reached.

There are many examples of rather ordinary communication contexts in which issues of meaning negotiation arise very naturally.

A simple and powerful example has been defined by Furnas et al. (1987) as the "vocabulary problem". Studying human-system interactions, Furnas et al. found that even in simple naming tasks individuals rarely agree *ex ante* on which word to use for referring to common objects or situations. They found that in such cases there are in general no perfect synonyms, and there is a low probability of *ex ante* lexical agreement between two different individuals – this difficulty of *ex ante* lexical agreement is also at the core of a popular computer interactive web game, the ESP game by von Ahn (2006).[1]

[1] In the ESP game (so called because it encourages players to "think like each other"), two players randomly matched through the web have to find a common (agreed) label for an image. The game has become a prototype for the concept of "Games with a purpose", since human participants'

As individuals do not agree *ex ante* on the lexical choice, and differences in their preferred one may actually mark subtle differences in the way they conceptualize the situation at hand, how do they converge on a sufficiently agreed lexicon during a conversation or other types of communicative interaction? Brennan and Clark (1996), in their analysis of the "vocabulary problem", have submitted that this happens through "conceptual pacts" – temporary agreements about how the referent is conceptualized. Once such a "pact" is reached, individuals can repeatedly and confidently refer to an object with the same term – which translates into the familiar phenomenon of lexical entrainment (Garrod and Anderson 1987; Brennan 1996; Pickering and Garrod 2004), i.e. the tendency of people to adopt the terms introduced by their interlocutor within a conversation.

Brennan and Clark notice some features of conceptual pacts which are worth reporting here. First of all, "conceptual pacts are established by speakers and addressees jointly" (Brennan and Clark 1996, 149). They are the result of an interactive process that may involve different rounds, lexical proposals and counterproposals, and may imply also disagreement. Furthermore, lexical pacts are specific to a given speaker-addressee pair. In other words they tend to reflect the specific relation between the two and the process through which an agreement has been reached – the same speaker may reach different pacts with different addressees. The emergence of conceptual pacts on the early stages of a conversation has been shown to be a good predictor of the overall cooperative success of communication (Reitter and Moore 2007; Nenkova et al. 2008).

Another interesting example of continued negotiation of concepts, where payoffs are not just semantic, comes from Andersson (1994). He investigates how different meanings of "nature" are used and argued for by different social, political and cultural groups. For example, he documents the tensions in the meaning of "natural forest" between forest owners, environmentalists and government officials and their power struggles to establish their preferred meaning. The outcome of the negotiations will have economic, environmental and legal consequences.

The use of vague predicates in communication Parikh (1994) provides another neat example of the ubiquitousness of meaning negotiation. When a vague predicate is asserted in communication, this often corresponds to a move that proposes to restrict the range of its possible values. As Barker (2002, 2013) suggests, by stating that "Harrison Ford is a tall actor" a speaker suggests that any actor taller than Harrison Ford is tall as well – if the addressee accepts this statement, all actors taller than Harrison Ford will be automatically annexed as tall to the common ground of the conversation. A parallel statement that "Tom Cruise is not a tall actor" would introduce a new restriction on the interval of tall actors, narrowing the range of admissible standards of tallness for actors. Of course, some of these statements may be rejected by an addressee – for example by rejecting the assertion that "Johnny Depp is tall" the addressee would signal her refusal to concede that the standard of tallness falls below the 1.80 m. limit.

---

playfulness is used to solve problems that are difficult to solve in automated ways – in this case image labeling (von Ahn 2006).

The fact that vague predicates are intrinsically underdetermined invites their renegotiation in the context of each specific conversation. Through negotiation, agents can reach an agreement that sufficiently restricts the vague area to satisfy the coordination needs of communication – or decide that they cannot agree. As such, meaning negotiation contributes to the flexibility of vague predicates, and makes them adaptable to different contexts.

Another interesting, and more subtle, case of meaning negotiation is related to indirect speech (Pinker et al. 2008). Indirect speech often reveals the presence of conflicting preferences in communication, and the need of communicating agents to negotiate through language the understanding of their mutual relations. Why should people often blur their communicative intents by allusive expressions or euphemisms? Pinker and coauthors suggests that, among other motivations, indirect speech reveals uncertainty about the intentions of the listener, and are often first moves in a series of language manoeuvers allowing to explore possible agreements without incurring the psychological (or sometimes material) cost of rejection. For example, a driver trying to bribe the policeman fining him might try a phrase such as "so maybe the best thing would be to take care of it here" (Pinker et al. 2008, 834), thus checking the honesty of the policeman without making an explicit offer that might lead to an accusation of bribery. The strategic nature of indirect speech is even more apparent when communication does not play only the role of transmitting information but also supports the negotiation of reciprocal relations between two persons (as in the case of an allusive sexual offer).

5.2 Negotiating in Conceptual Spaces

Conceptual spaces (Gärdenfors 2000) provide a very natural framework for modelling meaning negotiation. In Warglien and Gärdenfors (2013) and Gärdenfors (2014) we develop an account of meaning as emerging from the interaction of different individual conceptual representations – as a "meeting of minds". By taking a radical departure from traditional semantics, we state that the meaning of an expression does not reside in the world or (solely) in the mental schemes of individual users, but rather emerges from the mappings between individual mental spaces that are established through communication. The fundamental role of a communicative act, in this view, is to try to bring about cognitive changes (van Benthem 2008) by affecting others' states of mind.

The "meeting of minds" framework is couched in a geometric view where concepts are represented as convex regions of conceptual spaces, and the emergence of meaning is modelled as resulting from the mutual convergence of the positioning of each agent in the "product space" of their mental representations.

A simple example of such a process, that can be used as a more general metaphor of the emergence of meaning from interaction, is the achievement of joint attention in children's pointing (Bates 1976; Brinck 2004; Gärdenfors and Warglien 2013). A meeting of minds occurs in pointing when child and mother perceive that, as a result

of an original directional gesture, they are aligning their focus of attention on the same point in the surrounding physical space. When this convergence happens and is mutually recognized (e.g. through mutual gazes), the child's picture of what he is pointing out to the mother agrees with his understanding of what she is attending to (the same for the mother), and a sort of communicative equilibrium point is established, which can be formally modelled as a fixed point in the mappings between the mental spaces of the interacting agents (Warglien and Gärdenfors 2013).

The emergence of meaning in linguistic communication can be seen as a sort of generalized pointing process, in which language is used to drive the other's mind in a desired direction in her own mental space. A formal analysis of such communication processes shows that convexity of concepts plays an important role in ensuring that a "meetings of minds" exist. Other features of linguistic communication further support the existence of such points and the possibility to reach them. For example, the fact that the lexicon can express the categorization of an underlying conceptual space allows the use of discrete language tokens to approximate fixed points, while pragmatic maxims of conversation à la Grice (1989) facilitate convergence to such points (Warglien and Gärdenfors 2013).

However, in Warglien and Gärdenfors's original formulation of interactive meaning, little attention has been paid to the role played by differences in individual preferences for a given conceptualization. For example, different lexical preferences applied to the same object may translate into different conceptualizations of that object. When two individuals referring to the same brick wall use "wall" and "barrier", they may categorize differently the same visual scene and express a different communicative intention, reflecting different preferences for the scene representation (they "point" to different conceptual entities). In this case, finding a mutual agreement may imply some lexical give and take through which a meaning negotiation happens.

In order to understand meaning negotiation, one needs to develop a notion of an individual commitment to some preferred representation – for example a given categorization of an object, a certain combination of quality features characterizing a product, or a representative example of a tall actor. In general, the nature of conceptual preferences can be purely cognitive – for example the result of the individual learning history. In other cases, it may reflect the value of a specific conceptualization in the light of broader utility considerations – for example, the interpretation of the prototypical quality of the object of exchange in a commercial contract (e.g. what "a workmanlike job" means in a construction contract) is subject to obvious conflict of interest between the two transacting parts. Furthermore, one needs to develop a notion of what makes it acceptable to diverge from such preferred representations, as well as of what will make divergence unacceptable – to the point that we might prefer an open disagreement.

A simple way to capture the essence of a meaning negotiation problem in conceptual spaces is to assume that individuals have preferred points in such space, and that there is a subjective cost in departing from a preferred point. For example, I may have my own preferred threshold for separating tall from non-tall actors in the

dimension of height. I may be willing to accept departures from such threshold for the sake of conversation, but the larger the divergence from my own standard, the larger my discomfort. Thus, the cost of divergence from my favorite point will be a function of distance from such point (distance in conceptual spaces will express some measure of dissimilarity). And beyond a certain point, the discomfort caused by such divergence will offset the advantages of keeping our conversation running – and disagreement will break out.

The idea can be expressed graphically in a simple way. Consider a two-dimensional conceptual space (e.g. the space of beer "strength", defined by a combination of bitterness and alcohol degree). Let us assume that two individuals mutually engaged in communication share the same two-dimensional conceptual space (later we will relax this assumption), but have two different points defining a prototypical "strong beer", respectively A and B. While both have an interest in developing their conversation, assume that the expected benefits of a conversational agreement are reduced by any deviation from their preferred point. If the cost of such deviations is an increasing function of distance from the ideal point, after a given distance there may be no further interest in agreeing with the other agent, and conflict will be preferred to concession. Thus, for each agent, the area of acceptable definitions of a strong beer would be a circle (Fig. 5.1) around respectively point A and B. The circumference of the circles delimit what each player can afford as the maximum acceptable distance from the prototype. The intersection of such circles will define a set of possible agreements – what both can accept as a definition of strong beer (Fig. 5.1a). Not all the points of the possible agreement set have the same status, though. Points that are outside the segment connecting A and B are in a strong sense (Pareto) inefficient: agents can improve their position without damaging the other by coming closer to such segment. Thus, the bold segment (a, b) will define the efficient agreement set, where agreements should be expected to fall (Fig. 5.1b).

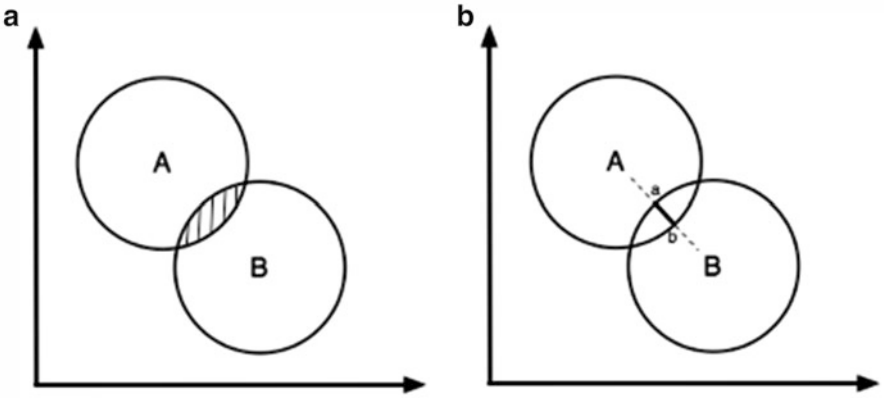

Fig. 5.1 (**a**) Acceptable agreements (**b**) Efficient agreements

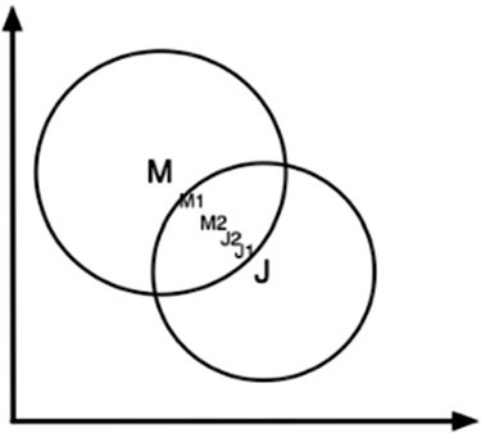

Fig. 5.2 Moves in the bargaining process

While this powerfully restricts the set of agreements one is expected to observe, it leaves still indetermined the problem of which agreement should prevail. Game theorists have provided a large repertoire of solution concepts for situations such as the one described above. One solution that most naturally fits communication contexts however pre-dates game theory – Zeuthen et al.'s (1930) approach to bilateral bargaining. Zeuthen's idea can be simply summarized by a key question: at each stage of a bargaining process, who should make a concession? The outcome of a bargaining problem will depend from the progressive contraction of the bargaining space as a result of subsequent concessions.

Imagine a situation like the one depicted in Fig. 5.2. The possible agreements set is not empty, thus a solution to the meaning negotiation problem may be expected. Mary and Joe have both made initial proposals M1 and J1 that fall within the agreement set, but leave room for further negotiation. For example, Mary could accept Joe's offer (after all it's in the acceptable set), but of course her own proposal is more convenient to her. Its not clear, though, that Joe could accept Mary's proposal, so Mary could be tempted to make a concession by proposing, for example, M2. Similarly, Joe has to decide whether to stay on his own proposal – at the risk of triggering a conflict with Mary – or make in turn a concession, say J2.

Zeuthen's key assumption is that at each turn Mary (Joe) will evaluate the situation, and assess what is the maximum acceptable risk that Joe (Mary) will prefer the conflict rather than accept M1 (J1). The intuition is that the one who has the lowest acceptable risk will concede. In other words, those that can afford a larger risk have an advantage. An important implication is that if agents are symmetric (same cost of conflict, same risk aversion) the solution will be an even split along the efficient agreements line. Otherwise, players enjoying a comparatively lower conflict cost and having higher propensity to risk will be more willing to engage in conflict and thus have higher bargaining power.

The idea that players more willing to face conflict have an advantage seems empirically reasonable also in the communication domain – and indeed experimental

data on how individuals negotiate a language can support such claim. For example, Selten and Warglien's (2007) experiment on the emergence of a common (artificial) lexicon in a two-person language game of referring shows that players signaling their stubborness in the early stages of the game have a decisive advantage in imposing their own preferred code for referring to a set of objects. Brennan's (1996) study on lexical entrainment in man-machine interaction shows that individuals having a conversation with a computer dialogue interface tend to concede to the lexicon of the (credibly stubborn) machine.

Under the further assumption that individuals act according to expected utility, Zeuthen's mechanism leads to a well-known solution concept in cooperative game theory, the Nash bargaining solution (Harsanyi 1956).[2] While we will not go into the details of such solution concept in this chapter, it is worth noting that the same topological properties (compactness and convexity) of conceptual spaces that support "meetings of minds" (Warglien and Gärdenfors 2013) support the existence and uniqueness of the Nash solution.

This view of meaning negotiation crucially depends on the fact that some intial representation is established for each agent, and that agents can locate their meanings in such space. It may be questioned that it is always possible to open up new dimensions in a negotiation. However, since concepts have an "open texture" (Porosität) (Waismann 1968), there is always some new aspect of a concept that has not been captured by the negotiation. Waismann argues that concepts (outside mathematics) can never be given a complete definition in terms of necessary and sufficient conditions: "[T]here will always remain a possibiility ... that we have not taken into account something or other that may be relevant to their usage. ... I shall never reach a point where my description will be completed" (1968, 121–122). Broader situations in which meaning has to be jointly elaborated through a process of search of relevant dimensions (see Egré 2013) may lay outside the scope of our use of "meaning negotiation".

5.3 Variations

Different assumptions may lead to slightly different ways of determining the equilibrium solution (Thomson 1994) – and the fact that in many cases language is discrete may lead anyway to just approximations of such solutions (Warglien and Gärdenfors 2013). Equilibira solutions are motivated by normative considerations.

[2]The Nash solution predicts that players will jointy maximize the product of their utilities. The Nash solution shoud not be confounded with the concept of Nash equilibrium. The Nash equilibrium and the Nash solution to the bargaining problem belong to two different families of game theoretic solutions, the former being a non-cooperative games solution concept, the latter a cooperative games one: see Osborne and Rubinstein (1994) for an accessible introduction to both. For example, the Nash solution assumes Pareto-efficiency as an axiom, while Nash equilibria can be non Pareto-efficient.

However, there are also other argumentation tools that may be involved in reaching an agreement or a partial agreement in a negotiation. In this section we present some tools of this kind.

Interesting implications can be derived by assuming some form of bounded rationality, that may limit the ability of agents to have a full rational control of all the space of representations. For example, many conversational phenomena appear to be driven by automatic processes rather than deliberation (Pickering and Garrod 2004). Also, it has been shown that the choice of the reference word in lexical comparison can alter the salience of properties of given objects affecting similarity judgments (Ortony et al. 1985). In these cases, the effects of changes (or manipulations) of representations can be considered, that can substantially affect the outcome of meaning negotiation.

5.3.1 Salience Manipulation

Consider a case such as is depicted in Fig. 5.3. The distance (conceptual dissimilarity) between the preferred points C and D is such that it overrides the benefits of cooperation, and lack of agreement should be expected. However, the distance between such points depends on how the two dimensions are weighted – something that will depend crucially from the salience attributed on each dimension (Gärdenfors 2000). Appropriate manipulation of the salience of different dimensions can modify the perception of dissimilarity between two agents (Ortony et al. 1985) and modify the distance between the two preferred points, thus facilitating the emergence of a possible agreement area (see Fig. 5.3). It is well known that salience effects can be manipulated in conversation and affect perceptions in automatic, hard-to-control ways (Taylor et al. 1979). For example, a speaker may exploit the priming effects of mentioning some words early in the conversation to make the dimensions associated to such words more relevant through entrainment, a mostly automatic process (Pickering and Garrod 2004).

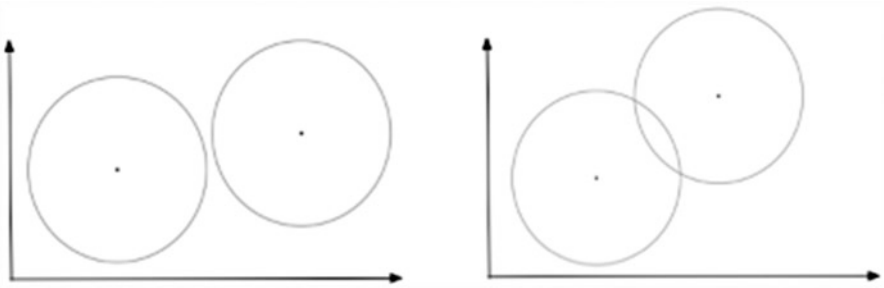

Fig. 5.3 Agreements are made possible as the weight of the abscissa is reduced

5.3.2 *Partial Agreements*

The use of semantically underdeterminate words is frequent in language (Ludlow to appear). While in some cases it can be related to simple reasons of economy, e.g. when a more determinate description is not needed, some level of indeterminacy may be related to the search for a partial agreement when a full one is not reachable. Typically, a partial agreement will consider only some dimensions of the problem, ignoring or deferring other ones. For example, as concepts can have multiple quality dimensions but adjectives typically represent only single or integral (non-separable) attributes (Gärdenfors 2000, 2014), a conversation focusing on a specific adjective will implicitly foster partial lexical pacts.

Once more, conceptual spaces suggest a natural analysis of such a phenomenon. Consider again two individuals with apparently incompatible conceptualizations, such that the circles representing their area of acceptable meanings do not overlap (Fig. 5.4). Despite the global incompatibility of their concepts, they still can agree on the projections of the circles on the single conceptual dimensions x or y (the x_c and y_c segments in Fig. 5.4). Thus, there is room for at least a partial agreement on each dimension. One classical example is the separation of disagreement on facts from disagreement on values, the former being often easier to solve (Perelman and Olbrechts-Tyteca 1958/1969). On a micro-level, conversational templates such as the proverbial "It doesn't matter whether a cat is white or black, as long as it catches mice" provide an obvious template for partial agreements.

Of course, which dimension prevails might create some advantage for one of the two speakers, which implies that there is ample room for different forms of dialogue manipulation (Van Eemeren and Grootendorst 2004; Van Eemeren 2010). For example, as the projection over a subset of conceptual dimensions can be considered as a degenerate case of salience manipulation (in which one or more

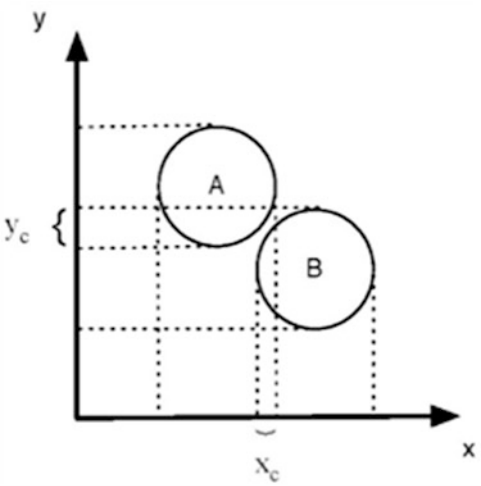

Fig. 5.4 Partial agreement areas on projections

dimensions have zero salience), all the conversational moves trying to focus on single properties (e.g. exploiting entrainment) can have the effect of facilitating a partial agreement on dimensions favorable to the speaker.

Partial agreements are also likely when representations of speakers have different dimensions, sharing only some of them. In that case, it appears as unavoidable that meaning negotiation will be performed on the shared dimensions, leaving others often implicit and thus leading to intrinsically underdetermined agreements.

While partial agreements emphasize dimensional reduction as a strategy to make an agreement possible, the symmetric manipulation, i.e. introducing new dimensions, is of course possible. Adding new dimensions might be motivated by the necessity to search for solutions in a broader space, but also respond to other strategic considerations, such as the need for an agent to move out of a negotiation space where he has a comparative disadvantage. We don't further elaborate this case here, alhough it is clearly relevant for meaning negotiation dynamics and it is a natural development of our approach.

5.3.3 Metaphoric Projection

Students of negotiation often stress the key role that metaphors play in the linguistic interaction that lead to negotiated agreements. But how do metaphors affect meaning negotiation itself? We suggest that metaphors play a key role in meaning negotiation by performing at a same time a selection process over dimensions and a modification of the similarity structure of the discourse domain.

Metaphors are commonly understood as mappings that transfer structure from a source domain to a target one. Such mappings act selectively on both the source and target domains – they select specific structural aspects of the source and mold the target according to such structural aspects – thus only those dimensions of the target which are compatible with the target are selected: "the lion Ulysses" emphasizes Ulysses' courage but hides his condition of a castaway in Ogiya. Thus metaphors act by orienting communication and selecting dimensions that may be more or less favorable to the speaker. By suggesting that a storm hit the financial markets, a bank manager can move the conversation away from dimensions pertaining to his own responsibilities and instead focus on dimensions over which he has no control, strengthening his position vis à vis his audience (Rocci 2009).

At the same time, metaphors shape the distance between different points in conceptual spaces by providing context for their interpretation – e.g. by providing contrast classes within which distance between elements is modified. This can be illustrated by a more complex example, provided by how the Falling Dominoes metaphor, dominating foreign policy in the 1950s and 1960s, created a representation that brought close to each several countries otherwise differing in terms of political and military issues; for example, it downplayed those aspects of the North Vietnamese position related to nationalism to emphasize ideological dimensions shared with other countries of the Communist block (McNamara et al. 2007).

This led for example to significantly downplay the strong distance between North Vietnam and China. The Falling Dominoes metaphoric blindfold forced alternative positions in foreign policy into a funnel that significantly narrowed disagreements over possible policies.

5.3.4 Non Cooperative Aspects of Meaning Negotiation

Until now we have assumed that in meaning negotiation agents will agree on a point which is in the efficient set of possible agreements – the standard assumption of cooperative games. However, it may be useful to remove this assumption in order to analyze the emergence of conversational failures. LiCalzi and Maagli (2013) have analyzed the problem of negotiating the categorization of conceptual spaces using the tools of non-cooperative game theory. The example of the negotiations of the meaning of "natural forest" from Andersson (1994) is a real life illustration of such a situation.

There are two agents and each one of them has a conceptual space that for analytic convenience is represented as a circle. Each agent categorizes the conceptual space in two convex partitions, Left and Right (the dividing line needs not be a diameter). Agents have an incentive to agree on the same categorization of the circle, but if the agreed partition is different from their preferred one, they incur a disutility proportional to the area they are "giving up" to the agreement – so each rational player strives to minimize her losses (Fig. 5.5).

This simple structure allows singling out two types of initial disagreement. One is the "focused" disagreement, where the lines dividing the conceptual spaces for each agent do not cross each other – thus the disagreement area is a convex region between the two lines (Fig. 5.6a) The other one is the "widespread disagreement", where the lines partitioning the conceptual spaces for each agent cross each other. This implies that the area of initial disagreement is not convex (Fig. 5.6b). As we shall see, these initial conditions of disagreement have important implications for the solution reached.

The case of focused disagreement is simpler and illustrates how the game works. It basically happens to be a zero sum game in which what is lost by one player is gained by the other one. A simple sequential process gives the first move in the

Fig. 5.5 Two partitions (from LiCalzi and Maagli 2013)

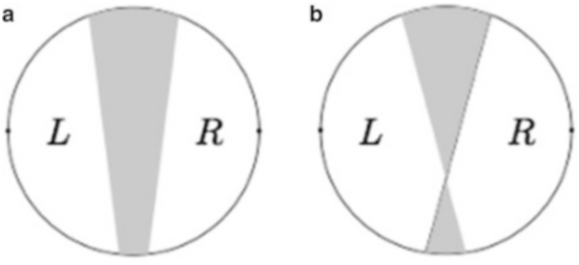

Fig. 5.6 Focused (**a**) and widespread (**b**) disagreement (from LiCalzi and Maagli 2013)

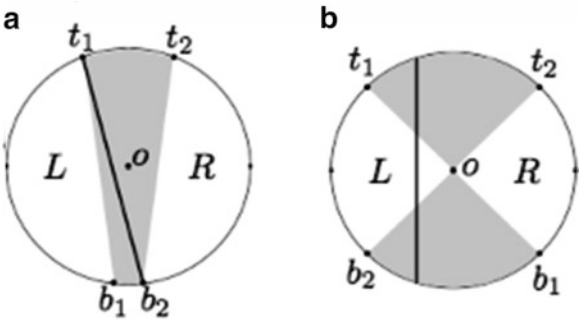

Fig. 5.7 Nash equilibria of the game for the focused (**a**) and widespread (**b**) disagreement (from LiCalzi and Maagli 2013)

game to one player, mimicking the possibility of the first speaker to create an anchor to the establishment of a "common ground". He or she can thus choose where to locate a first point on the circle's circumference. The second player can only pick a second point that will determine how the circle is partitioned. If both players are trying to minimize their disutility, they will both concede nothing of their original partition, and will stick to one extreme or their dividing line. The result is shown in Fig. 5.7a: the player who controls the longest arc (e.g. the one who can control the communciation "agenda") loses less than the other (the shaded area represnts the initial disagrameent area).

The case of widespread disagreement is more complex (Fig. 5.7b). To see it assume (with no loss of generality) that the two dividing lines (at the beginning of the game) go through the center of the circle. It would seem reasonable that at the end fo the game the dividing line passes through the center, providing an efficient solution. Unfortunately, the Nash equilibrum of the game, given the "stubborness" of players in minimizing individually their disutility, leads to a communication failure: both players lose some of the potential area of consensus, generating a solution that is inefficient – that could be improved if both acted in a more collaborative way. Thus, a simple conversation game can show the emergence to a sort of "conversational dilemma", the failure of communication to preserve the pre-existing consensus.

5.4 Discussion

Despite its pervasiveness, meaning negotiation is still a rather under-analyzed phenomenon. We can only speculate on why it is so. Three reasons stand as rather natural. First of all, meaning negotiation presupposes a view of language in which semantic underdetermination plays an important role – a view certainly in contrast with the central tenets of classical semantics. Furthermore, it presupposes a view of meaning as (at least to some extent) a social, interactive phenomenon, once more violating the strong view of meaning as fundamentally independent from the communicative interaction of speakers.

Finally, while the pragmatic tradition concedes a significant role to communicative interaction, it still relies heavily on the assumption that language in use is a collaborative enterprise (Clark 1996), leaving in the shadow aspects related to conflict between communicating agents.

We claim that a geometric approach to meaning (Gärdenfors 2000, 2014; Warglien and Gärdenfors 2013) is well equipped to deal with such "anomalous" features of meaning negotiation. It allows to explicitly represent underdeterminateness in terms of regions of a meaning space. It allows to naturally represent the interactive nature of meaning via mappings between different individual meaning spaces. And it can represent conflicting preferences for meanings as different locations in such spaces.

In this chapter we have shortly sketched the outline of a geometric view of meaning negotiation. We have shown that such view can inherit important structural elements from game theoretic models of bargaining – in particular, for the case where the protagonists have overlapping negotiation regions, we have emphasized a parallel to the Nash solution in cooperative game theory. When acceptable solutions regions of the protagonists are disjoint, we have presented several types of processes: changes in the salience of dimensions, dimensional projections and metaphorical space transformations. None of the latter processes are motivated by normative or rationality considerations, but presented as argumentation tools that we believe are used in actual situations of conceptual disagreement.

Acknowledgements Massimo Warglien recognizes financial support by the *MatheMACS* project, funded by the European Commission under the 7th Framework Programme Grant #318723. Peter Gärdenfors thanks the Swedish Research Council for support to the Linneaus environment *Thinking in Time: Cognition, Communication and Learning*.

References

Andersson, T. (1994). *Conceptual polemics: Dialectic studies of concept formation* (Lund University cognitive studies 27). Lund: Lund University.

Barker, C. (2002). The dynamics of vagueness. *Linguistics and Philosophy, 25*(1), 1–36.

Barker, C. (2013). Negotiating taste. *Inquiry, 56*(2–3), 240–257.

Bates, E. (1976). *Language and context: The acquisition of pragmatics*. New York: Academic.

Brennan, S. (1996). Lexical entrainment in spontaneous dialog. In *Proceedings, 1996 international symposium on spoken dialogue, ISSD-96* (pp. 41–44). Philadelphia, PA.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1482–1493.

Brinck, I. (2004). The pragmatics of imperative and declarative pointing. *Cognitive Science Quarterly, 3*(4), 429–446.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*(2), 259–294.

Egré, P. (1913). What's in a planet? In M. Aloni, M. Franke, & F. Roelofsen (Eds.), *The dynamic, inquisitive and visionary life of phi? phi and Diamond-phi, A Festtschift for J. Groenendijk, M. Stokhof and F. Veltman* (pp. 74–82), ILLC, 2013.

Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Communications of the ACM, 30*(11), 964–971.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press.

Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. Cambridge, MA: MIT Press.

Gärdenfors, P., & Warglien, M. (2013). The development of semantic space for pointing and verbal communication. In J. Hudson, U. Magnusson, & C. Paradis (Eds.), *Conceptual spaces and the construal of spatial meaning: Empirical evidence from human communication* (pp. 29–42). Cambridge: Cambridge University Press.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27*(2), 181–218.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Harsanyi, J. C. (1956). Approaches to the bargaining problem before and after the theory of games: A critical discussion of Zeuthen's, Hicks', and Nash's theories. *Econometrica, 24*(2), 144–157.

Larson, R. K., & Ludlow, P. (1993). Interpreted logical forms. *Synthese, 95*(3), 305–355.

LiCalzi, M., & Maagli, N. (2013). *Bargaining over a common conceptual space*, manuscript. Working Papers 30, Department of Management, Università Ca' Foscari Venezia.

Ludlow. (to appear). *The dynamic lexicon*, manuscript.

McNamara, R. S., Biersteker, R. S. M., Blight, J., Brigham, R. K., Thomas, J., Blight, J., Brigham, R. K., Biersteker, T. J., & Schandler, C. H. (2007). *Argument without end: In search of answers to the Vietnam tragedy*. New York: Public Affairs.

Nenkova, A., Gravano, A., & Hirschberg, J. (2008). High frequency word entrainment in spoken dialogue. In *Proceedings of the 46th annual meeting of the association for computational linguistics on human language technologies: Short papers* (pp. 169–172). Stroudsburg: Association for Computational Linguistics.

Ortony, A., Vondruska, R. J., Foss, M. A., & Jones, L. E. (1985). Salience, similes, and the asymmetry of similarity. *Journal of Memory and Language, 24*(5), 569–594.

Osborne, M., & Rubinstein, A. (1994). *A course in game theory*. Cambridge, MA: MIT Press.

Parikh, R. (1994). Vagueness and utility: The semantics of common nouns. *Linguistics and Philosophy, 17*, 521–535.

Perelman, C., & Olbrechts-Tyteca, L. (1969). *The new rhetoric: A treatise on argumentation*. Notre Dame: University of Notre Dame Press.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(02), 169–190.

Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences, 105*(3), 833–838.

Reitter, D., & Moore, J. D. (2007). Predicting success in dialogue. In *Proceedings of the 45th annual meeting of the association for computational linguistics* (pp. 808–815). Prague: Association for Computational Linguistics.

Rocci, A. (2009). Maneuvering with voices. In F. H. Van Eemeren (Ed.), *Examining argumentation in context: Fifteen studies on strategic maneuvering*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.
Selten, R., & Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. *Proceedings of the National Academy of Sciences, 104*(18), 7361–7366.
Stalnaker, R. (1999). *Context and content: Essays on intentionality in speech and thought*. Oxford: Oxford University Press.
Taylor, S. E., Crocker, J., Fisle, S. T., Sprinzen, M., & Winkler, J. D. (1979). The generalizability of salience effects. *Journal of Personality and Social Psychology, 37*(3), 357–368.
Thomson, W. (1994). Cooperative models of bargaining. In R. J. Aumann & S. Hart (Eds.), *Handbook of game theory with economic applications* (Vol. 2, pp. 1237–1284). Amsterdam: Elsevier.
Van Benthem, J. (2008). "Games that make sense": Logic, language, and multi-agent interaction. In K. R. Apt & R. Van Roij (Eds.), *New perspectives on games and interaction* (pp. 197–210). Amsterdam: Amsterdam University Press.
Van Eemeren, F. H. (2010). *Strategic manoevering in argumentative discourse*. Amsterdam: John Benjamins Publishing Company.
Van Eemeren, F. H., & Grootendorst, R. (2004). *A systematic theory of argumentation: The pragma-dialectical approach*. Cambridge: Cambridge University Press.
Von Ahn, L. (2006). Games with a purpose. *Computer, 39*(6), 92–94.
Waismann, F. (1968). Verifiability. In A. G. N. Flew (Ed.), *Logic and language* (pp. 117–144). Oxford: Blackwell.
Warglien, M., & Gärdenfors, P. (2013). Semantics, conceptual spaces, and the meeting of minds. *Synthese, 190*(12), 2165–2193.
Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.
Zeuthen, F., Zeuthen, F. E., & Wiggs, K. I. (1930). *Problems of monopoly and economic warfare*. London: Routledge & Sons.

Part III
Computing Meanings

Chapter 6
How to Talk to Each Other via Computers: Semantic Interoperability as Conceptual Imitation

Simon Scheider and Werner Kuhn

Abstract What exactly does interoperability mean in the context of information science? Which entities are supposed to interoperate, how can they interoperate, and when can we say they are interoperating? This question, crucial to assessing the benefit of semantic technology and information ontologies, has been understood so far primarily in terms of standardization, alignment and translation of languages. In this article, we argue for a pragmatic paradigm of interoperability understood in terms of *conversation* and *reconstruction*. Based on examples from geographic information and land cover classification, we argue that semantic heterogeneity is to a large extent a problem of multiple perspectives. It therefore needs to be addressed not by standardization and alignment, but by *articulation and reconstruction of perspectives*. Reconstruction needs to be grounded in shared operations. What needs to be standardized is therefore not the perspective on a concept, but the procedure to arrive at different perspectives. We propose *conceptual imitation* as a synthetic learning approach, and *conceptual spaces* as a constructive basis. Based on conceptual imitation, information provider and user concepts can be checked for perspectival correspondence.

6.1 Semantic Interoperability: From Convention to Conversation

Metadata and semantic technology can be seen as cornerstones of a working information society.[1] Increasingly, they are being recognized as such. The dominant search engines on the Web, Google, Yahoo, and Bing, have incorporated light

[1] Compare the role of metadata in Gray et al. (2005).

S. Scheider (✉) • W. Kuhn
Institut für Geoinformatik, Westfälische Wilhelms-Universität Münster, Heisenbergstraße 2, D-48149 Münster, Germany
e-mail: simonscheider@web.de

F. Zenker, P. Gärdenfors (eds.), *Applications of Conceptual Spaces*,
Synthese Library 359, DOI 10.1007/978-3-319-15021-5_6

weight ontologies and present facts semantically related to a search result.[2] The Linked Data Web helps libraries, governments and museums provide seamless access to their archival data, which have remained information silos for decades (Hyvönen 2012). Bioinformatics,[3] Environmental Science (Madin et al. 2007), as well as Geoinformatics (Janowicz et al. 2012) have realized some time ago that their data sets of diverse origin and uncontrolled authorship require semantic metadata for successful sharing.

Semantic technology draws on methods from ontology engineering, computational inference, information retrieval, and machine learning (Stuckenschmidt and van Harmelen 2003). The purported goal is to help computer systems, and, indirectly, people, interoperate across cultural, domain and application boundaries. However, despite this development, there still remains a central blind spot: *What exactly does interoperation mean?* Which entities are supposed to interoperate, how can they interoperate, and when can we say that they are interoperating? Skeptics have not ceased to question the usefulness of explicit knowledge representations, and for very good reasons (Shirky 2009). Semantic interoperability as a research goal remains largely underexposed. Central underlying concepts, such as the relationship of information and interoperation have remained obscure and incomprehensible. This leaves the benefit of semantic technology blurry, and negatively impacts its credibility.

In related work on data semantics (Janowicz and Hitzler 2012; Janowicz 2012; Scheider 2012), it has become apparent that semantic engineering of digital information may be regarded as a matter of dynamic *machine mediated communication* between data providers and users, with data (as well as semantic meta-data) being the explicit top of a pyramid of implicit acts of interpretation, observation and construction.

In this article, we argue for a pragmatic shift in semantic interoperability. The proposed paradigm of conversation and reconstruction holds that data has meaning in a pragmatic communication context which implies a *perspective*. The problem is that this perspective is lost, in one way or another, under the conditions of digital communication. Semantic engineering therefore needs to support the communication and reconstruction of this perspective, not necessarily its standardization or translation. The goal is to support users in comparing perspectives and in judging their usefulness for their own (often diverging) goals. We propose to use *conceptual spaces* for this purpose. In this article, we make the case for a new paradigm which is meant to provoke future research into this direction.

In the remainder, we will first motivate the new paradigm by analyzing examples of land cover categories. We will then suggest how information sharing in human-machine-human conversations can be based on *conceptual imitation* and conceptual spaces. The last section demonstrates how conceptual imitation can be used to reconstruct and compare land cover classes.

[2] http://schema.org/

[3] http://zacharyvoase.com/2009/11/02/bioinformatics-semweb/

6.2 Semantic Interoperability Revisited

A shift from convention to conversation in semantic technology mirrors the historic shift in linguistics and philosophy of logic which has led to the development of *speech act theory* (Austin 1953). In his seminal paper, Grice (1975) argued that the debate between formalists and informalists in philosophical logic may be overcome if formal logics could be made to pay adequate attention to the conditions that govern *conversation*. Conversation depends on perspective. For example, the sentence "he is in the grip of a vice" may refer equally well to one's bad habit or to the fact that part of one's body is caught in a tool. The problem is that any standardized, conventional meaning of the word "vice" may be just not what was meant by the speaker, as this meaning shifts with speech situations. A system of inference which is based on conventional or default word meanings is therefore bound to fail in human communication. Insofar as data is a vehicle for human communication, we may be confronted with a similar situation.

6.2.1 Semantic Interoperability as a Problem of Multiple Perspectives

The first example we discuss is taken from an early article on interoperability of geospatial data (Harvey et al. 1999; Riedemann and Kuhn 1999). Suppose we had to represent a sportsground like that depicted on the aereal photograph in Fig. 6.1 in a cartographic map for the purpose of noise abatement planning. Our goal is to identify objects on the sportsground that emit noise.

Which information resource about the sportsground will provide us the required information? Figures 6.2 and 6.3 show two very different kinds of maps. In the cadastral map (Fig. 6.2), the most prominent feature, the track and soccer field area, is not shown. Similarly, the tennis courts are left out. This can be explained if we recall that a cadastral map is about land parcels and ownerships, and that the distinction between a soccer field and its surrounding area is not one of ownership. Google maps seems to stick to a similar cadastral world view in Fig. 6.4.

In contrast, the sportsground itself is represented in the cadastral map since it can be distinguished from its surroundings precisely based on ownership. The topographic map in Fig. 6.3, on the other hand, shows the track area and the tennis courts but leaves out the soccer field. This can also be explained if we recall that a topographic map is about ground surface features. There is a distinction in surface texture between the elliptic track area or the tennis courts on one hand and the lawn on the other.

However, since our goal is to identify sources of noise, we are interested in identifying the soccer field. Surface texture does not allow to distinguish it from its embedding lawn, and so the maps discussed so far are not interoperable with our purpose. Rather, we need to switch perspective on the sportsground and regard

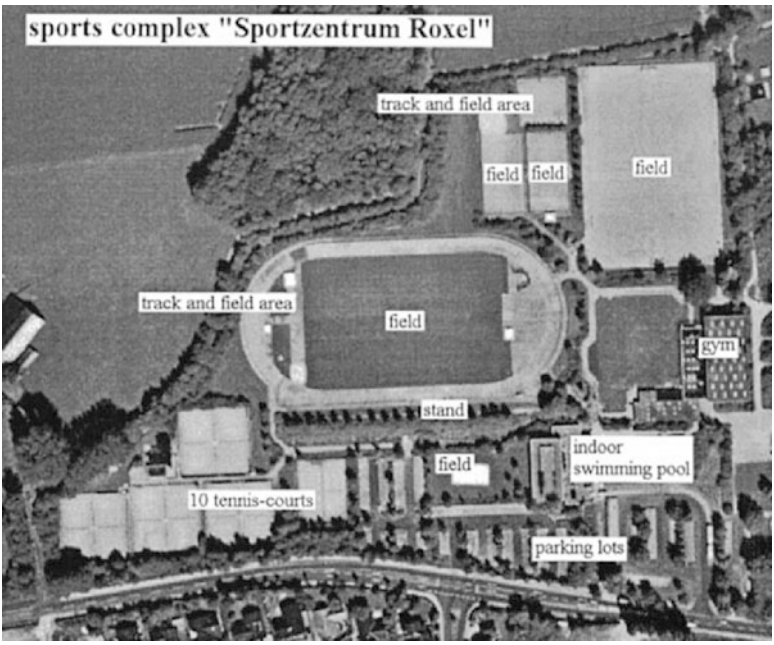

Fig. 6.1 Sportsground Roxel near Münster in an aereal photograph (Taken from Harvey et al. (1999))

Fig. 6.2 Cadastral map (ALK) of the Sportsground Roxel (Harvey et al. 1999)

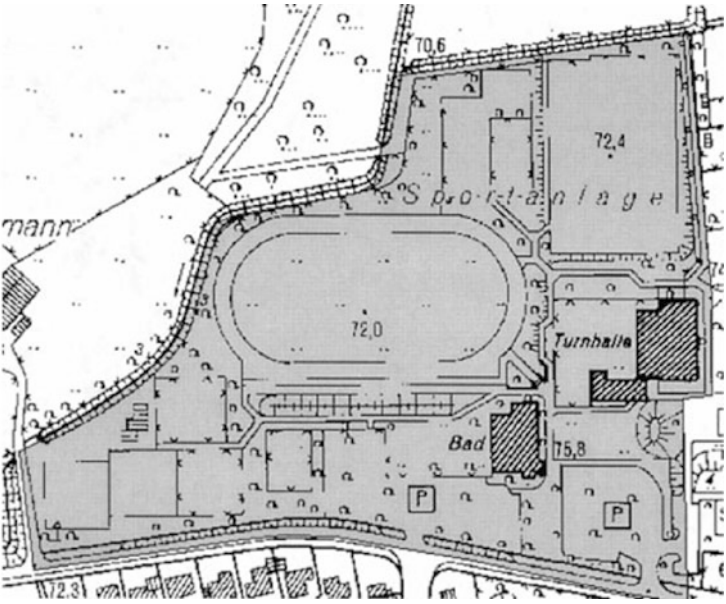

Fig. 6.3 Topographic map (DGK) of the Sportsground Roxel (Harvey et al. 1999)

Fig. 6.4 The sportsground on Google maps

it in terms of *social affordances* (Scarantino 2003). The latter are conventionally established dispositions to play a game, indicated by linear signs on the lawn.[4] Under the new perspective, some aspects of the surface, such as the signs, become relevant, while others, such as the texture, become irrelevant.

It seems that there is large variability in mapping a single area, to the extent that even individual geographic phenomena, not only their semantic categories, appear and disappear depending on the kind of abstraction one is imposing on an observed environment. Moreover, it is not the environment, but rather the purpose, and with it the kind of observation and abstraction employed, which allows to distinguish the cadastral from the topographic and the game perspective. These perspectives are based on different underlying concepts, namely ownership, surface texture and gaming affordance.

Many geoscientific categories have an intrinsic multi-perspectival nature. They are situated, i.e., their interpretation depends on space and time, and sometimes on a group of researchers (Brodaric and Gahegan 2007). This frequently causes interoperability problems (Brodaric 2007). A further example is the assessment of the environmental impact of land use on our climate. *IPCC*[5] *land cover classes* serve to quantify land use change based on a transition matrix such as in Fig. 6.5.

	Forest land	Grass-land	Crop-land	Wet-lands	Settle-ments	Other land
Forest land	FF	FG	FC	(FW)	FS	FO
Grass-land	GF	GG	GC	(GW)	GS	OS
Crop-land	CF	CG	CC	(CW)	CS	CO
Wet-lands	(WF)	WG	WC	WW	WS	WO
Settle-ments	(SF)	(SG)	(SC)	(SW)	SS	(SO)
Other land	OF	OG	OC	(OW)	OS	OO

Fig. 6.5 Land cover categories of the IPCC in a transition matrix used to determine land use change for climate impact assessment. Each element denotes a transition class (Source: Global Forest Observation Initiative (GFOI))

[4] Similar to traffic locomotion affordances (Scheider and Kuhn 2010).

[5] International Panel on Climate Change, http://www.ipcc.ch/

The matrix is used to classify land area change (with every matrix element standing for a type of transition) in an attempt to estimate the impact of land use on our climate.

The problem is that the IPCC land cover classes allow for a *large degree of freedom of interpretation* and *do not distinguish incommensurable perspectives*. The continuous transition of surface texture from forest to grassland allows for arbitrary category boundaries. This causes an explosion of more than 800 locally adopted forest definitions (Lund 2012), each of which has a specific ecological (and political) context. Furthermore, and more importantly, some classes such as cropland and forest draw on incommensurable perspectives on a land surface. From an ontological standpoint, a *crop* is not a kind of plant, *it is a role played by a plant in the course of human cultivation*. Oil palm plantations, for instance, can be considered croplands or forests, depending on whether one takes a vegetation or a cultivation perspective (Lund 2006). This makes the two categories end up on orthogonal conceptual dimensions, instead of being mutually exclusive and uni-dimensional, as required by the IPCC (Lund 2006).

The examples demonstrate that semantic interoperability is to a significant extent a problem of multiple perspectives. This frequently causes fruitless debates about how phenomena should best be represented "in general", as documented, e.g., by the object-field dualism (Couclelis 1992). The existence of multiple purpose dependent perspectives on a domain suggests a pragmatic approach to semantics, as proposed, e.g., by Brodaric (2007) and Couclelis (2010).

Multiperspectivity is a basic trait of human culture, cognition and perception. The multiple perspectives underlying language were put sharply into focus by Quine[6] and Wittgenstein (2003). Analogously, human perception was recognized early by Gestalt psychologists as being *multi-stable*, i.e., allowing to switch between different geometric interpretations of the same scene (Köhler 1992). This is demonstrated, e.g., by the Necker cube or the face/vase illusion in Fig. 6.6. It seems therefore that we need to pay closer attention to the different pragmatic techniques[7] that deal with perspective. In how far do current semantic engineering approaches take account of this?

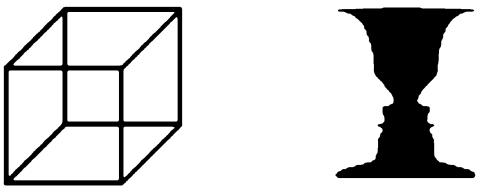

Fig. 6.6 Necker cube and face/vase illusion can be perceived in terms of two contradicting 3-D interpretations by the subconscious visual routines (Lehar 2003)

[6] Quine based his famous arguments of indeterminacy of theory (Quine 1951) as well as reference (Quine 1974) on multiperspectivity.

[7] Better captured by the German term "Kulturtechniken".

6.2.2 Paradigms of Semantic Interoperability

What is it that makes semantic interoperability a challenging problem? In essence, we seem to underestimate multiperspectivity by assuming that it can be solved through standardization or translation.

- The problem cannot be reduced to *labeling*, i.e., to establishing standard terms or term mappings for given concepts. A labeling problem is a *multi-term/single perspective* problem. It implies that the inventory of concepts, i.e., individuals and categories, is fixed, and interoperability problems arise only because we use different names for the same concepts.
- Interoperability is also not a *translation problem*. Translation is a standard approach in the Semantic Web and ontology mapping (compare chapter 2.4 in Stuckenschmidt and van Harmelen 2003). In the translation approach, we assume that the sets of concepts underlying different datasets may be different but can be mutually translated based on a *shared ontology*. The latter is a theory where each concept is definable so that a new concept has a translation in the ontology (Stuckenschmidt and van Harmelen 2003). For example, a new concept may be a hyponym of an existing one and thus may be related to it by subsumption.[8] This approach requires, however, that a shared ontology exists.
- In addition, interoperability cannot be assured by *standardization*. Standardization of names solves only labeling problems. And standardization of concepts (in terms of a formal theory) requires a common perspective which serves as a denominator for all the different perspectives. This, however, is precisely what can not be expected under multiperspectivity.

Table 6.1 lists different paradigms that semantic engineering has gone through so far, together with a new one, called reconstruction and conversation. The paradigms are ordered by their tolerance with respect to semantic heterogeneity. While holistic standardization tries to resolve heterogeneity by applying a technical standard (an example would be an ISO[9] standard), top-level ontology alignment seeks to avoid heterogeneity by sticking only to a top-level standard, as proposed, e.g., in Masolo et al. (2003). Pluralist peer-to-peer translation, in contrast, does not require an ontology standard. It acknowledges heterogeneity and at the same time tries to mitigate its negative effects by establishing translations between ontologies based on similarity (Euzenat and Shvaiko 2007). The ability to translate between ontologies implies at least an overlap in central concepts (Stuckenschmidt and van Harmelen 2003). Concept similarity, however, is only a necessary and not a sufficient condition of concept identity. What if concepts are similar but at the same time untranslatable, because they correspond to different perspectives on the same matter, as in the case of landcover classes?

[8] Compare Chapter 6.1 in Stuckenschmidt and van Harmelen (2003).

[9] http://www.iso.org

Table 6.1 Paradigms of semantic interoperability

Paradigm	Main idea	Strategy	Required means	Basic assumption
HOLISTIC STANDARDIZATION	Term-meaning standard	Heterogeneity resolution	Ability to subscribe to a standard	Term-meaning can be standardized
TOP-LEVEL ONTOLOGY ALIGNMENT	Alignment with core standard	Heterogeneity avoidance	Ability to align with core standards	Core term-meaning can be standardized
PLURALIST PEER-TO-PEER TRANSLATION	Term similarity and translation	Heterogeneity mitigation	Ability to translate between similar terms	Term-meanings are comparable because concepts overlap
RECONSTRUCTION AND CONVERSATION	Term-meaning regeneration	Heterogeneity articulation	Ability to reconstruct concepts and to act on information	Concepts can be reconstructed and term-meanings can be communicated

Semantic interoperability rather seems to be a *multi-term multi-perspective* problem. This means that perspectives vary in fundamental ways with respect to their ontological commitment which makes them to some degree *untranslatable* (Quine 2001). That is, each dataset comes with concepts which are not present in another perspective, since they have entirely different origins. For example, think about translating ownership terms into vegetation surface terms on the same sportsground.

For the purpose of comparison, however, it is still possible to expose different concept origins. We suggest therefore that interoperability should be based on *reconstruction* and *conversation*. Reconstruction involves knowing how information was obtained, i.e., which observations and interpretations where performed and how abstractions were generated. It enables users to understand differences and to regenerate conceptualizations. Conversation validates this knowledge in a *peer-to-peer* fashion, i.e., with respect to a certain data producer. This is in analogy to Grice's conversationally fixed meaning. Following Janowicz and Hitzler (2012) and Janowicz (2012), the strategy therefore should not be to resolve, avoid or mitigate heterogeneity, but to articulate it. This implies that semantic differences need to be understood and be focused on, not leveled. For this purpose, term-meanings need to be (re)generated and communicated, not standardized, aligned or translated, see Fig. 6.7.

The principal problem with the existing paradigms is their weakness in *distinguishing perspectives*, i.e., in detecting basic conceptual differences and thus translatability in the first place. On what grounds do we assume that there is a shared ontology? And vice versa, how could we possibly know that no such ontology exists? It seems that the current paradigms do not provide good answers to these two questions. Also, their focus on ontology alignment and translation risks technically enforcing some kind of unification on superficial grounds.

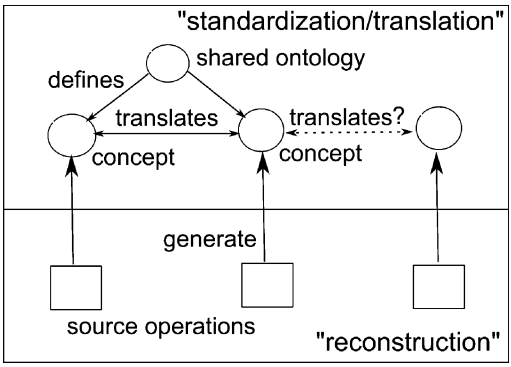

Fig. 6.7 Reconstruction vs. standardization/translation paradigm. In the reconstruction paradigm, the impossibility of translations can be assessed based on reconstructing concepts with source operations

One should note that the paradigms in Table 6.1 are not meant to substitute each other. Conceptual standards and translations are neither useless nor wrong, merely insufficient to detect multiple perspectives. Instead of standardizing perspectives, we suggest to *standardize procedures to arrive at different perspectives*.

6.3 Purpose, Design and Sharing of Information in a Machine Context

In order to address perspectivity by pragmatic means, we suggest to take a pragmatic view on information as such. Janich (2006) has argued that one of the most far-reaching errors of modern science is the fairy tale of "information as a natural object" which is supposed to exist independently of cultural techniques.[10] The problem with this idea is that it tends to blur exactly those origins of information that we need to expose. Namely, that information is *designed for a purpose*, and that the *sharing of information* is simply a function of this design and the capabilities involved.

Design makes information a product of (communication) culture, not nature (Couclelis 2010). Therefore, the distinction between pragmatics and semantics of information, as well as between situated and non-situated concepts (Brodaric 2007), seems spurious. The sharing of digital information is a matter of its design in a human-machine-human conversation.

[10]The history of this fundamental misunderstanding can be traced back to Morris' naturalized semiotic process and Shannon and Weaver's mechanistic information theory, and can be currently studied in terms of modern nonsense about "information" allegedly being "transferred and understood by machines, computers, and DNA molecules" (Janich 2006).

6.3.1 *Information in Human-Machine-Human Conversations*

The problem is not that machines cannot communicate, but that humans misunderstand each other when communicating via machines (Scheider 2012). At first glance, digital machines have greatly increased the efficiency, speed and range of human interaction. They have successfully substituted humans in searching, filtering, transforming and validating information, to the extent that their role appears to be negligible. The Semantic Web can be seen as an effort to lift this to the level of meanings.

However, the latter view neglects that *information* is basically a *derivative of human speech acts*, and that the success of any information-processing machine therefore needs to be measured with respect to human capacities of speech (compare Janich 2006: Chapter 5). Information appears if people inform other people about something. The acts involved turn data into information. This remains true even if human speech is technically encoded, transmitted and extended by computers and technical sensors (Fig. 6.8). The limits of modern information technology seem therefore less defined by the technical efficiency of symbol processing, but by the *substitution of human speech acts by equivalent technical capacities*. Regarding the latter challenge, modern information technology turns out to be surprisingly weak.

For example, there is a fundamental problem regarding the encoding of reference (Scheider 2012), which is illustrated, e.g., by the debate about the meaning of URI[11] references in the Semantic Web (Booth 2006; Halpin 2013). What does a URI refer to, over and above the Internet address to which it can be resolved? Digital information is basically a product of encoding *acts of reference* to a domain experienced by an encoder (Fig. 6.8). In terms of visual indices to jointly perceived scenes, acts of reference play a central role in learning and sharing a language

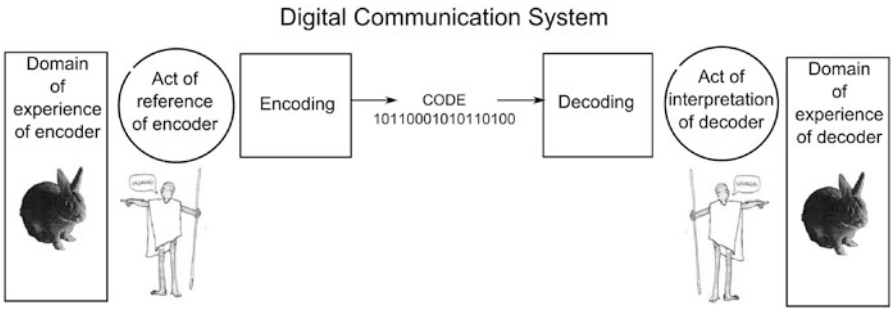

Fig. 6.8 The problem of reference in human-machine-human conversation. Acts of reference are encoded in a digital form, which makes it hard for a decoder to interpret the code in terms of his or her own domain of experience

[11] Uniform resource identifier.

(Tomasello 1999; Quine 1974). However, a language standard for encoding resource reference, such as RDF,[12] does not ensure that the decoder understands and shares referents.

The role of information technology is to substitute speech acts by equivalent capacities in order to increase the range and efficiency of human speech. The challenge is to *talk to each other via computers*, even if communicators are not present in the same scene (Janich 2006). What is needed for this purpose?

We know that the efficiency and range of human speech can be effectively extended beyond natural boundaries using signs. This is demonstrated, for example, by a traffic light which prompts a driver to stop and move even in the absence of a traffic policeman. Similarly, the turn signal on a vehicle indicates a turn without the driver being visible (Kamlah and Lorenzen 1996). However, in order to understand such signs, knowledge about the underlying human capacities is essential, as they are the methodological ground for understanding. Correspondingly, traffic lights and turn signals are meaningless in the absence of knowledge about prompted stopping and turning actions.

In a similar sense, data sets can be regarded as manifest speech acts (Scheider 2012). Speech acts *prompt an intended kind of reaction* from a listener. It is this reaction that the listener needs to understand. For example, in Fig. 6.8, the encoder draws the attention of the decoder to a rabbit in his or her domain of experience. More precisely, the encoder prompts a focusing of attention with a symbol like "Gavagai" (reusing the famous example from Quine 2001). Furthermore, the encoder makes a statement about this rabbit, e.g., that it was observed at a location at a time. If this statement is digitally encoded, then we obtain *data*. However, only if the drawing of attention was successful is reference to the rabbit shared among agents, and the data set then becomes meaningful for the decoder. Otherwise, the decoder remains unsure about what was meant. Thus, we need to ask how we can make users *learn the reactions intended by the providers of a data set*.

6.3.2 The Law of Uphill Analysis and Downhill Invention

The challenge we face reflects a more general one in AI and robotics, first formulated by Valentino Braitenberg, a cybernetician, in his famous book about "vehicles" (Braitenberg 1986). Braitenberg presented a neat collection of very simple construction plans for technical devices made from analog circuitry ("vehicles") that are able to generate a complex set of human-like behaviour, such as love, fear, thought and optimism. For example, by a simple analog circuit and a light intensity sensor, a vehicle can be made to move in such a way as to avoid light sources.

[12]Resource Description Framework, http://www.w3.org/RDF/

Braitenberg's thought experiment shows that it is easy to create an interesting repertoire of complex behaviour based only on simple processes with feedback.[13] This is "downhill invention". However, from the outside, without opening the black box, an observer of the vehicles' behaviour has almost no chance of correctly guessing how they were built. The content of the black box may look as if it were beyond the grasp of human reason, just as love and thought continuously appear to human analysts. This is "uphill analysis". There are several reasons why analysis is harder than synthesis. First, many different potential mechanisms could realize the same behaviour. And second, induction is harder than deduction because one lacks the constructive basis, i.e., "one has to search for the way, whereas in deduction one follows a straightforward path" (Braitenberg 1986, p. 20).

Braitenberg suggests that *analysts* of a mechanism therefore tend to overestimate its complexity because they only have access to complex behaviour, not to the constructive basis. They first *describe* complex behaviour (which is complex), and then try to guess how it may be generated. However, once analysts know which basis is the correct one, things start to become easy, and they become *synthesists*.

This insight applies also to semantic interoperability. If we take Braitenberg's law seriously, then *synthetic learning needs to take on an indispensable role in the learning of meanings*.[14] This explains why it is hard to handle semantic interoperability under the familiar paradigms. Current approaches to data integration are either based on "black box" descriptions of concepts (*ontologies*), or on *machine learning* of semantic concepts. The latter is an automatization of guessing a function based purely on observations of external behaviour (Hastie et al. 2001). The former are a specification of purported behavioural and conceptual constraints in a top-down fashion (Guarino 1998). However, both approaches are basically analytic, not synthetic. That is, they rely on descriptions of concept behaviour and do not involve any information about how a concept was generated and on which constructive basis.[15] This may be one reason why semantic interoperability is hard with analysis tools, while human language learning is almost effortless for children.

What does it take, then, to interoperate with information? First, "interoperation" should be taken literally, i.e., it consists of speech acts that intertwine and prompt reactions, even if increasingly large parts of these acts are performed by machines. And second, the prompted kinds of reactions include those that need to be learned in a synthetic fashion. Synthetic learning is what makes information a *designed* product. And this is the tough part of interoperability, because it requires the correct constructive basis.

[13] For a proposal how central spatial concepts can be based on Braitenberg vehicles, see Both et al. (2013).

[14] Compare also the arguments given in Chapter 3 of Scheider (2012).

[15] Machine learning is analytic in the sense that it prescribes a constructive basis (e.g., in terms of a vector calculus in support vector machines (Hastie et al. 2001)) or automatically selects it based on observed behaviour (as in Bayesian model selection).

6.3.3 *Synthetic Learning Requires Imitation*

How do we learn in a synthetic rather than analytic fashion?

If we follow somebody in synthesizing something, then we *imitate* this person based on our own capacities. For example, if a child is taught how to build a castle, it learns by imitating the construction (even if the result may look different, and even if its competences are slightly different). In the same sense, *sharing meaning requires imitation*, and as such seems to be the fundamental mechanism for knowledge transmission, as argued by Tomasello (1999). For example, the robot on the hand left side of Fig. 6.9 learns the concept of "holding something in front" by recomputing the meaning of this notion in terms of its own body frame of reference. That is, it does two things simultaneously: it observes a speech act behaviour of its opposite and then recomputes a concept based on its own sensori-motor system. Note that the latter system largely constrains the constructive basis. The system basically consists in a vector calculus grounded in body postures. The grounding leaves no question which constructive basis to choose, and thus learning is synthetic.

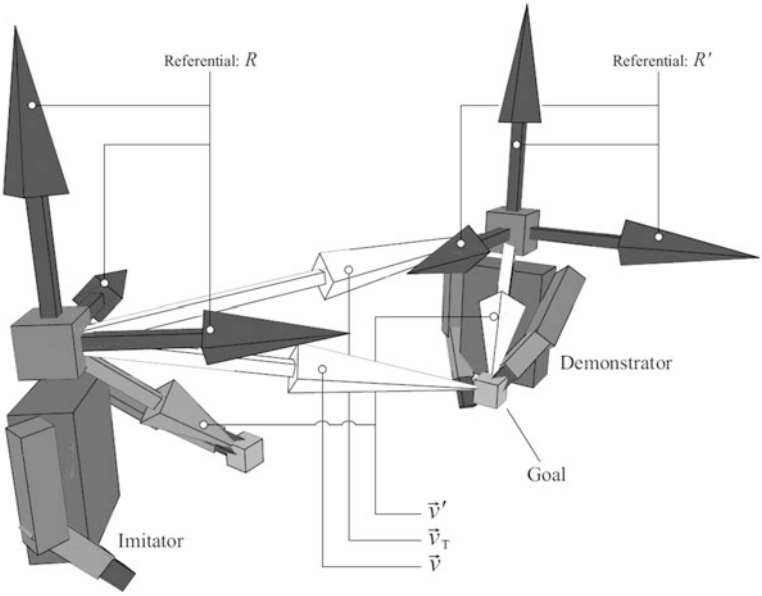

Fig. 6.9 Robot imitation learning of "holding something in front" based on recomputing it with respect to body reference frames (Sauser and Billard 2005). The imitator robot needs to translate vector v' relative to the demonstrator referential frame into a corresponding vector of its own referential frame based on vectors v,v' and v_T

6.3.4 Synthetic Learning Needs to Be Grounded

The insight that synthetic learning and imitation needs to proceed from a *common embodied ground* is central in embodied AI and robotics (Steels 2008). We argue that this is also the case for synthetic learning in data semantics, because the grounding disambiguates the constructive basis for a concept (Scheider 2012).

Luc Steels has demonstrated that robots can construct simple perceptual distinctions and share them via *discrimination games* (Steels 1997). This is an example of synthetic learning of concepts. Based on this, robots can exploit social mechanisms of language learning (Steels 2003) which allow them to establish term-meanings and their own languages via so-called *language games* (Steels 2002).

Grounded symbol systems exist also in human communities and play an immensely useful role in technology, namely in the form of *reference systems*. There are reference systems for measurements (measurement scales), for locations (geodetic coordinate systems) as well as for time (calendars), allowing us to refer to phenomena in a very precise manner. It is important to understand, however, that not every formal symbol system is grounded (Harnad 1990). In particular, grounding is not implied by formal ontologies (Scheider 2012). For many ontological concepts, reference systems and groundings are currently lacking. Kuhn (2003) therefore proposed to build new *(semantic) reference systems* for arbitrary kinds of concepts, and Scheider (2012) discussed how grounding of such systems might be done.

As it turns out, there is a large variety of constructive calculi as well as primitive perceptual operations that may be used for this purpose. Beule and Steels (2005), for instance, have proposed Fluid Construction Grammar as a formalism. The latter follows the idea of Johnson Laird's procedural semantics (Johnson-Laird 1997), where the meaning of a phrase is taken to be the execution of a (perceptual) program. For humans, a rich set of perceptual operations is available for any reproducible phenomenon that can be perceived, including surfaces, actions and action potentials (Scheider 2012, Chapter 5). Measurement procedures and technical sensors simply increase the range of human perception (Scheider and Stasch 2015). From the viewpoint of logic, constructive calculi range from sets of constructive rules (as in intuitionist logic) (Lorenzen 1955), over recursive functions in higher-order logic (HOL) (as in functional programming) (Nipkow et al. 2002), to the formation of logical definitions in ordinary first-order logic (FOL) (Scheider 2012, Chapter 4). They may, in particular, involve geometric constructions (Scheider and Kuhn 2011). Gärdenfors' *conceptual spaces* are a special but very essential case, where concepts are constructed as convex regions (Gärdenfors 2000). A *grounding level* therefore is not a static thing, and it is essential that it has been established, either by convention and standardization or by practice, in order to start synthetic learning.

6.4 Interoperability Through Conceptual Imitation

In this section, we discuss tools that support conceptual imitation. We propose to use Gärdenfors' conceptual spaces (Gärdenfors 2000) as a convenient constructive basis for synthetic learning. We regard a conceptual space as a *grounded multidimensional space* equipped with vector space operators, following Adams and Raubal (2009), where the grounding can be secured either in terms of conventionalized human perception or in terms of measurement. Concepts are generated in this space by all operations that generate convex regions. Conceptual imitation requires to establish conceptual spaces and to describe and reconstruct concepts in terms of them. Reconstructed concepts can then be compared, and it can be assessed whether they are interoperable or not. We demonstrate this on the basis of land surface category examples discussed in Sect. 6.2.

6.4.1 Conceptual Imitation Tools and Conceptual Spaces

Which tools can be used to support agents in synthetic learning? In principle, this can be all tools for grounded concept construction. Most generally, we describe concepts in terms of *predicates*, i.e., in terms of concept symbols (F) with slots for instances (a) which can be used to state that the former symbols apply to the latter ($F(a)$). Thus we need to look for tools that support the reconstruction of predicates in some grounded language. A reference theory (Scheider 2012) is a formal system of symbols that comes with a fixed operational interpretation of the following kind (compare also the left hand side of Fig. 6.10 and Table 6.2):

1. A set of *grounded instances* (the domain). These may consist of *foci of attention* of an observer (Scheider 2012), or of *technical foci* of sensors according to their discrimination unit (Frank 2009; Scheider and Stasch 2015).[16] They may also consist of conventionally established *objects*. Furthermore, predicates of interest (see below) may become instances at a higher level of abstraction, standing for classes of instances on the lower level (Scheider 2012).
2. A set of *grounded primitive predicates*, i.e., symbols with slots for instances whose meaning is fixed by standard procedures, as described below.
3. For each primitive predicate, a *known instantiation procedure*, which may be embodied and therefore not part of the formalism, and which allows to *decide* whether the predicate applies to grounded instances (i.e., it allows to instantiate the predicate) or not. Examples may be measurable sensor properties such as altitude and temperature, but may also involve perceptual distinctions of humans such as the discrimination of objects and their properties.

[16]The "instantaneous field of view" (IFoV) of a satellite is an example for the latter.

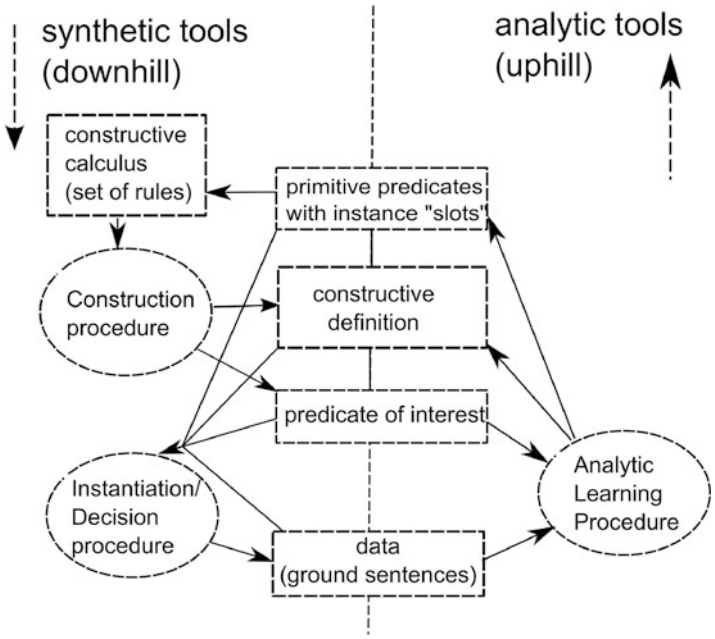

Fig. 6.10 Kinds of conceptual imitation tools and their role in learning. Synthetic tools start with a constructive basis, which consists of primitives and a calculus as well as an instantiation procedure for generating data, while analytic tools try to guess the constructive basis starting from the analysis of data

Table 6.2 Synthetic conceptual imitation tools

Imitation tool category	Grounded FOL example	Conceptual space example
GROUNDED INSTANCE	Focus of attention/sensor range/perceptual phenomenon	IFoV of some satellite
GROUNDED PRIMITIVE PREDICATE	Measurable or perceptual qualities	Ground altitude/tree height
CONSTRUCTIVE CALCULUS	Logical syntax/algebra	Vector calculus
CONSTRUCTIVE PROCEDURE	Formation algorithm	Convex region generator
CONSTRUCTIVE DEFINITION	Definiens	Convex region
PREDICATE OF INTEREST	Definiendum	Mountain/forest
INSTANTIATION PROCEDURE	Predicate satisfaction inference	Point-in-polygon test

4. A *constructive calculus*[17] on predicates, i.e., a set of syntactic rules which allow to generate (define) new predicates in terms of primitive predicates. Examples are the syntactic rules of a logic which may be used to generate definitions

[17] According to Lorenzen, a calculus is a set of rules used to generate "figures from other figures" (Lorenzen 1955).

of new predicates, or geometrical operations which may be used to construct new regions in a conceptual space. A particular construction is described by a *constructive procedure* or a *constructive definition* (definiens). The constructive calculus also involves semantic rules which prescribe how a defined predicate must be interpreted into instances.

5. A set of *predicates of interest* (definiendum) generated by constructive procedures.
6. An *instantiation procedure* for all predicates of interest, i.e., a sequence of inference steps that allows to decide whether these predicates apply to grounded instances, or not. The instantiation procedure is not (always) identical with the constructive procedure. For example, the syntactic pattern of a FOL definition does not imply a decision procedure. Only *decidable languages* have a decision procedure for every possible predicate of interest. If definitions are recursive, then inference procedures need to include a fixpoint operator. Note also that the instantiation procedure for predicates of interest may be distinct from that for primitives. Once the primitive predicates are determined, for example based on measurements, instantiation may proceed purely syntactically for defined predicates.
7. Instantiation procedures give rise to a set of *ground sentences*[18] about instances. They allow to compute whether a predicate (primitive or not) applies to instances, or not. Ground sentences are considered *data*.

Table 6.2 lists examples for imitation tools, drawn from FOL and conceptual spaces. It turns out that conceptual spaces are simple and straightforward examples of conceptual imitation tools. We illustrate this based on the categories *mountain* or *forest*, which can be defined in terms of convex regions in a space with dimensions based on remote sensors (compare also Sect. 6.4.3 and Adams and Janowicz 2011): Grounded instances are spatio-temporal granules (*instantaneous fields of view* (IFoV)), which correspond to pixels in a satellite image. Primitive predicates include ground altitudes and tree heights, which correspond to particular values on some measurement scale. Measurement scales correspond to single dimensions of a conceptual space, and combinations of values of different dimensions correspond to points in this space. The constructive calculus is a vector calculus, which allows to form algebraic expressions over points. Constructive procedures are algorithms for generating convex regions in this space, such as polytopes. Constructive definitions are definitions of convex regions, e.g., lists of points for polytopes. Predicates of interest are particular convex regions which account for a concept, e.g., mountain or forest polytopes. And instantiation procedures for predicates of interest are point-in-polygon tests, i.e., algorithms which determine whether a certain point lies within some region that defines a concept. For a more precise formulation of conceptual space operators, see Adams and Janowicz (2011), Adams and Raubal (2009), and for further illustrations regarding this example, see Sect. 6.4.3.

[18] These are sentences without variables.

All of the tools discussed above need to be learned in order to use them in conceptual imitation. Some tools, such as the constructive definition of a polytope, may be easily communicated, while others, such as the measurement procedure underlying tree height, may be more difficult to acquire.

Note that we assume the operations that allow *instantiating* primitive predicates to grounded instances to exist outside the reference theory, and thus outside of any computer. In the example above, these are measurement procedures of a satellite, but they could also be human observations. The requirement is not that these operations are *computable*, but that they are *shared and repeatable* in an inter-subjective way. The acknowledgment of decision procedures of different origin (embodied/analog/non-deterministic vs. algorithmic/digital) distinguishes our approach from ordinary logic. We also assume that predicates of interest come with a decision procedure. In our example, these are point-in-polygon tests for polytopes. In a more general setting, which may be based on the flexible FOL syntax, the ordinary FOL syntactic calculus needs to be either restricted, such as in the Semantic Web, or handled with some care to ensure that there is a decision procedure for each constructed predicate.[19]

Besides such synthetic tools, one may also use analysis tools, e.g., machine learning (Fig. 6.10). However, the latter come with a bias in their construction calculus (Mitchell 1980) and they are analytic, i.e., they lack information about the constructive basis. This leaves learners with the problem of choosing a constructive basis on their own, or of delegating the problem to some standard algorithm.

6.4.2 Measuring Perspectival Correspondence

Once a foreign concept has been learned synthetically, i.e., is reconstructed in terms of conceptual imitation, it becomes possible to measure the extent to which it corresponds to a concept in question or rather belongs to a different perspective. In principle, one can distinguish several levels of perspectival correspondence of concepts, depending on whether they share primitive predicates (with underlying operations) and correspond in terms of construction procedures (and constructive definitions), and whether they correspond to each other synthetically or analytically:

1. *Cocomputable*: The concepts are equivalent in the sense that they are constructed by the same procedure in terms of a calculus of deterministic primitive predicates.

[19]Note that we do not require a decision procedure for the entire constructive calculus, only for the predicates of interest. This allows the use of unrestricted FOL or HOL as the most flexible syntactic standard, but comes at the price of caring about the computation of decisions on a case-to-case basis.

2. *Codescendant*: The concepts are similar in the sense that they are constructed by the same non-deterministic predicates. This corresponds to drawing samples from the same random process.
3. *Comparable*: The concepts draw on the same primitive predicates, but may correspond to different procedures. This allows *translating* them into each other.
4. *PartiallyComparable*: Primitive predicates overlap. This allows projecting concepts onto common dimensions, but not translating them, since there are missing parts.
5. *Incomparable*: The concepts draw on different primitive predicates.
6. *Coincident*: Concepts apply to the same grounded instances.
7. *Overlapping*: Concepts overlap with respect to grounded instances.

Inside a conceptual space, concept similarity can be measured based on spatial proximity of concepts (Gärdenfors 2000). This is only possible if concepts are at least *comparable* or *partially comparable*; only in this case can we project concepts onto common dimensions. Our classification scheme for correspondence, however, covers also the case of incomparable concepts, where concepts are orthogonal to each other. In this case, similarity is not meaningfully captured by distance. However, there may still be equivalences on the level of types of phenomena that different conceptual spaces are covering, such as altitude, speed or temperature.[20]

Note that correspondences are not mutually exclusive. For example, some concepts may be coincident and incomparable, such as the concepts of heat and red, and may then be called *proxies*. Also, every codescendant is also comparable. Some relations may be refined gradually, such as Comparable, or Overlapping (the latter in terms of precision and recall). We discuss examples of these concept relations in the following.

6.4.3 Conceptual Imitation of Land Cover Categories

We illustrate the use of conceptual imitations by reconstructing land cover categories, as illustrated in Baglatzi and Kuhn (2013) and Adams and Janowicz (2011).

Suppose we have a vector space of *spatially continuous relief properties*, such as proposed by Adams and Janowicz (2011). Dimensions of this space may be altitude and relief measures as well as location, as depicted in Fig. 6.11. Each vector of this space represents a particular measurement in each of the metric dimensions at some measured focus in space (and time).

An elevation that is called a "mountain" in Scotland, such as Ben Lomond (974 m), would hardly be called the same in Asia. A relief space can be used for teaching and distinguishing local mountain concepts, for example *English mountains* and *Asian mountains*, and thus different perspectives on the concept

[20]This aspect of similarity is based on experiential equivalence and is not discussed in this article.

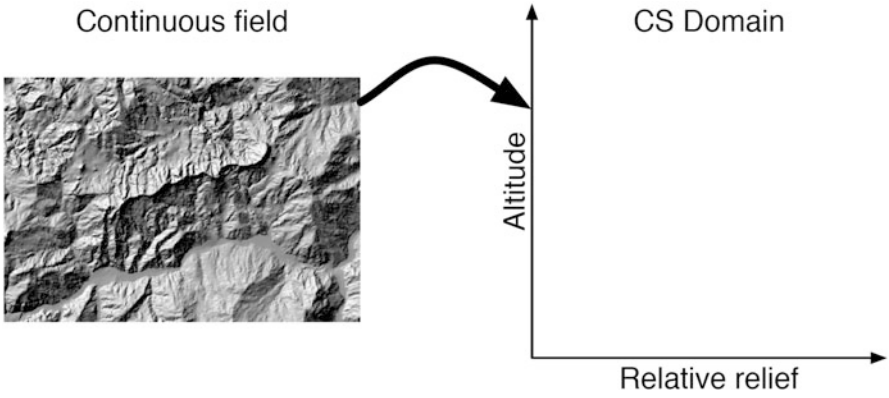

Fig. 6.11 A conceptual space of continuous altitude measures (Adams and Janowicz 2011). A continuous field of remote sensor pixels is mapped into a conceptual space of altitude and relief measures

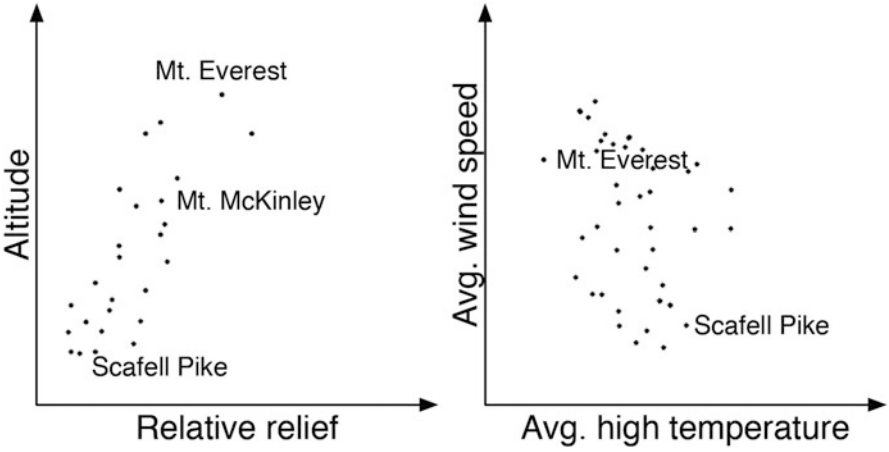

Fig. 6.12 Grounded mountain instances and their location in two conceptual spaces (Adams and Janowicz 2011)

mountain. The relief space comes with a set of primitive predicates, namely the set of values of relief measurements. There is an unambiguous decision procedure for each value in terms of the well known relief measurement procedures. The set of grounded instances is simply the finite sample of measurements taken. These instances can be projected into a set of vectors in Fig. 6.12. Note that measurement instances cannot be identified with points in this conceptual space, because the projection need not be injective, so that different measurements map into the same point. The vector calculus can be used to *define concepts as convex regions* in this space. The latter function as predicates of interest. Convex regions for the concept

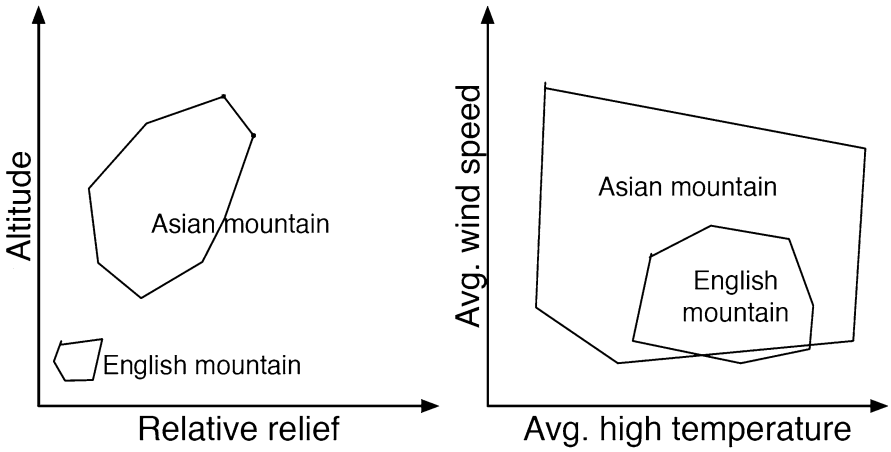

Fig. 6.13 Local mountain concepts synthetically learned in terms of convex regions (Adams and Janowicz 2011)

"mountain" can be efficiently computed based on convex polytopes (see Fig. 6.13). Point-in-polygon tests decide whether some instance lies in a polytop, and thus whether it falls under a specific mountain concept, or not.

In this way, one can find out to what extend English and Asian mountain concepts are different (compare Fig. 6.13); concepts are *non-overlapping*, yet defined in terms of the same grounding level, and thus *comparable and translatable*. In a similar way, one could define the meaning of hydrological object classes such as pond, lagoon and lake as regions in a space of size, naturalness, and marshiness, following the ideas of Mark (1993).

In analogical fashion, one may learn the IPCC land cover classes *cropland* and *forest*, as introduced in Fig. 6.5. The multitude of local forest definitions, as listed in Lund (2012), may be mapped into a conceptual vector space, provided observation procedures are established for the so called "Marrakesh" variables as dimensions. The latter include tree height, crown cover, and minimum area. They also imply an operational definition of the concept "tree", which needs to be established first, as well as expectation values of measured values for the future (Lund 2006). In such a conceptual space, one can discover that several different forest concepts overlap, see Fig. 6.14. According to our scheme, they are therefore *comparable*.

With the construction of the concept cropland, the situation is different. One first needs to ground a separate primitive predicate, e.g. *Cultivated*, that captures whether the land cover is used for growing crops (Lund 2006), i.e., whether it is an object of human cultivation. This predicate is conceptually orthogonal to the Marrakesh variables, which are not based on perceiving agricultural actions. Thus forest and *Cultivated* are not mutually exclusive.[21] However, since we now know

[21]This can be gleened from the fact that Lund (2006) proposes a decision tree which enforces mutual exclusiveness of cropland and forest by defining cropland based on cultivation as well as the logical complement of forest.

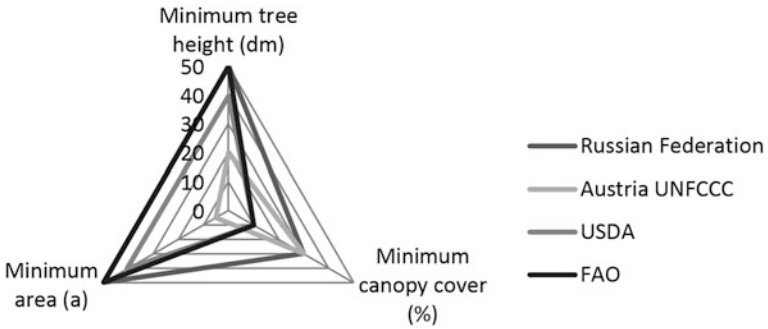

Fig. 6.14 Local forest definitions projected as triangles into a 3 dimensional vector space of "Marrakesh" variables (thresholds were taken from Lund (2012))

that *Cultivated* is defined on an entirely different operational basis than forest, we can conclude that both classes are based on incommensurable perspectives, they are *incomparable*. It may or may not be the case that forests tend to be non-cultivated. If not, the two concepts are also *overlapping*.

6.5 Conclusion

In this article, we have suggested that interoperability should be defined as *correspondence* of conceptual perspectives of communicating agents. Correspondence can be measured in terms of the constructive bases of the concepts in question. A data set of a provider is interoperable with respect to a user if the underlying user and provider perspectives correspond to each other in this sense. The pragmatic paradigm of conversation and reconstruction suggests that interoperability does not require to resolve, avoid or mitigate heterogeneity, but to articulate it. For this purpose, concepts need to be imitated, not standardized, aligned or translated. Correspondingly, digital information should be viewed as an abstraction of human speech acts that prompt imitations. Such imitations include synthetic learning. According to Braitenberg's law, synthetic learning involves knowledge about the constructive basis used to generate concepts, in contrast to analysis which is based solely on imitating concept behavior. We suggested therefore that imitation of semantic concepts should be grounded, and we proposed to do so based on conceptual spaces. This approach allows users to compare their intended concept with respect to a data provider concept, and thus to justifiably assess *the possibility of translations*. We illustrated the approach by conceptual imitation and assessment of correspondence of land cover classes.

Since this article is an outline of a pragmatic paradigm of interoperability, several questions remain open to future research. First, which grounding levels including calculi and primitive operations other than conceptual spaces are useful

for conceptual imitation? Second, how can conceptual imitation be technically realized as computer dialogues, such that users and providers are supported in their dialogue with the machine. And third, how can correspondence of grounded user and provider concepts be computed?

Acknowledgements A draft of the ideas in this article was presented at the EarthScienceOntolog session-3 at 10/11/2012.[22] Research was funded by the International Research Training Group on Semantic Integration of Geospatial Information (DFG GRK 1498), and by the research fellowship grant DFG SCHE 1796/1-1. We thank Helen Couclelis, Benjamin Adams, Krzysztof Janowicz and MUSIL[23] for discussions that helped shape this article.

References

Adams, B., & Janowicz, K. (2011). Constructing geo-ontologies by reification of observation data. In *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems, GIS '11*, Chicago (pp. 309–318). New York: ACM.

Adams, B., & Raubal, M. (2009). A metric conceptual space algebra. In K. Hornsby, C. Claramunt, M. Denis, & G. Ligozat (Eds.), *Proceedings of spatial information theory, 9th international conference, COSIT 2009*, Aber Wrac'h, September 21–25, 2009 (pp. 51–68). Berlin: Springer.

Austin, J. (1953). How to talk. Some simple ways. In *Proceedings of the Aristotelian society*, Bedford Square, London (Vol. 53, pp. 227–246).

Baglatzi, A., & Kuhn, W. (2013). On the formulation of conceptual spaces for land cover classification systems. In *Geographic information science at the heart of Europe* (pp. 173–188). Cham/New York: Springer.

Beule, J. D., & Steels, L. (2005). Hierarchy in fluid construction grammars. In *KI 2005*, Koblenz (pp. 1–15).

Booth, D. (2006). URIs and the Myth of Resource Identity. http://www.ibiblio.org/hhalpin/irw2006/dbooth.pdf.

Both, A., Kuhn, W., & Duckham, M. (2013). Spatiotemporal Braitenberg vehicles. In *21st SIGSPATIAL international conference on advances in geographic information systems, SIGSPATIAL 2013*, Orlando, November 5–8, 2013 (pp. 74–83). ACM.

Braitenberg, V. (1986). *Vehicles. Experiments in synthetic psychology*. Cambridge: MIT.

Brodaric, B. (2007). Geo-pragmatics for the geospatial semantic web. *Transactions in GIS, 11*(3), 453–477.

Brodaric, B., & Gahegan, M. (2007). Experiments to examine the situated nature of geoscientific concepts. *Spatial Cognition & Computation, 7*(1), 61–95.

Couclelis, H. (1992). People manipulate objects (but cultivate fields): Beyond the raster-vector debate in GIS. In A. Frank, I. Campari, & U. Formentini (Eds.), *Spatio-temporal reasoning* (pp. 65–77). Berlin/Heidelberg: Springer.

Couclelis, H. (2010). Ontologies of geographic information. *International Journal of Geographical Information Science, 24*(12), 1785–1809.

Euzenat, J., & Shvaiko, P. (2007). *Ontology matching*. New York/Secaucus: Springer.

Frank, A. (2009). Scale is introduced in spatial datasets by observation processes. In *Spatial data quality. From process to decision (6th ISSDQ 2009)*, St. John's (pp. 17–29). CRC Press.

[22] http://ontolog.cim3.net/cgi-bin/wiki.pl?EarthScienceOntolog

[23] musil.uni-muenster.de

Gärdenfors, P. (2000). Conceptual spaces – The geometry of thought. Cambridge: MIT.
Gray, J., Liu, D. T., Nieto-Santisteban, M., Szalay, A., DeWitt, D. J., & Heber, G. (2005). Scientific data management in the coming decade. *ACM SIGMOD Record, 34*(4), 34–41.
Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. 3, pp. 41–58). San Diego: Academic.
Guarino, N. (1998). Formal ontology and information systems. In N. Guarino (Ed.), *Formal ontology in information systems: Proceedings of the first international conference (FOIS '98)*, Trento, June 6–8 (Frontiers in artificial intelligence and applications, Vol. 46). Amsterdam: IOS-Press [u.a.].
Halpin, H. (2013). *Social semantics – The search for meaning on the web* (Semantic web and beyond, Vol. 13). New York: Springer.
Harnad, S. (1990). The symbol grounding problem. *Physica D, 42*, 335–346.
Harvey, F., Kuhn, W., Pundt, H., Bishr, Y., & Riedemann, C. (1999). Semantic interoperability: A central issue for sharing geographic information. *The Annals of Regional Science, 33*,213–232.
Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.
Hyvönen, E. (2012). *Publishing and using cultural heritage linked data on the semantic web* (Synthesis lectures on the semantic web). San Rafael: Morgan & Claypool Publishers.
Janich, P. (2006). *Was ist Information? Kritik einer Legende*. Frankfurt a. M.: Suhrkamp.
Janowicz, K. (2012). Observation-driven geo-ontology engineering. *Transactions in GIS, 16*(3), 351–374.
Janowicz, K., & Hitzler, P. (2012). The digital earth as knowledge engine. *Semantic Web Journal, 3*(3), 213–221.
Janowicz, K., Scheider, S., Pehle, T., & Hart, G. (2012). Geospatial semantics and linked spatiotemporal data – Past, present, and future. *Semantic Web, 3*(4), 321–332.
Johnson-Laird, P. (1997). Procedural semantics. *Cognition, 5*, 189–214.
Kamlah, W., & Lorenzen, P. (1996). *Logische Propädeutik. Vorschule des vernünftigen Redens* (3rd ed.). Stuttgart/Weimar: J.B. Metzler.
Köhler, W. (1992). *Gestalt psychology. An introduction to new concepts in modern psychology*. New York: Liveright.
Kuhn, W. (2003). Semantic reference systems. *International Journal of Geographical Information Science, 17* (5), 405–409.
Lehar, S. (2003). *The world in your head. A gestalt view of the mechanism of conscious experience*. Mahwah/London: Lawrence Erlbaum Associates.
Lorenzen, P. (1955). *Einführung in die operative Logik und Mathematik*. Berlin: Springer.
Lund, H. G. (2006). Guide for classifying lands for greenhouse gas inventories. *Journal of Forestry, 104*(4), 211–216.
Lund, H. (2012). Definitions of forest, deforestation, reforestation and afforestation. Technical report, Forest Information Services, Gainesville. http://home.comcast.net/%7Egyde/DEFpaper.htm.
Madin, J. S., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., & Villa, F. (2007). An ontology for describing and synthesizing ecological observation data. *Ecological Informatics, 2*(3), 279–296.
Mark, D. M. (1993). Toward a theoretical framework for geographic entity types. In A. U. Frank & I. Campari (Eds.), *Spatial information theory a theoretical basis for GIS* (Lecture notes in computer science, Vol. 716, pp. 270–283). Berlin/Heidelberg: Springer.
Masolo, C., Borgo, S., Gangemi, A., Guarino, N., & Oltramari, A. (2003). *Wonderweb deliverable d18: Ontology library*. Trento.
Mitchell, T. (1980). *The need for biases in learning generalizations* (Cbm-tr 5-110). Rutgers University.
Nipkow, T., Wenzel, M., Paulson, L. C. (2002). *Isabelle/HOL: A proof assistant for higher-order logic*. Berlin/Heidelberg: Springer.
Quine, W. (1951). Two dogmas of empiricism. *The Philosophical Review, 60*, 20–43.

Quine, W. (1974). *The roots of reference*. La Salle: Open Court Publishing.
Quine, W. (2001). *Word and object* (24th ed.). Cambridge: MIT.
Riedemann, C., & Kuhn, W. (1999). What are sports grounds? Or: Why semantics requires interoperability. In A. Vckovski, K. E. Brassel, & H. J. Schek (Eds.), *Interoperating geographic information systems* (Lecture notes in computer science, Vol. 1580, pp. 217–229). Berlin/Heidelberg: Springer.
Sauser, E. L., & Billard, A. G. (2005). View sensitive cells as a neural basis for the representation of others in a self-centered frame of reference. In *3rd international symposium on imitation in animals & artifacts*, Hatfield (pp. 119–127).
Scarantino, A. (2003). Affordances explained. *Philosophy of Science, 70*, 949–961.
Scheider, S. (2012). *Grounding geographic information in perceptual operations* (Frontiers in artificial intelligence and applications, Vol. 244). Amsterdam: IOS Press.
Scheider, S., & Kuhn, W. (2010). Affordance-based categorization of road network data using a grounded theory of channel networks. *International Journal of Geographical Information Science, 24*(8), 1249–1267.
Scheider, S., & Kuhn, W. (2011). Finite relativist geometry grounded in perceptual operations. In M. Egenhofer, N. Giudice, R. Morath, & M. Worboys (Eds.), *Spatial information theory: 10th international conference, COSIT 2011*, Belfast (Lecture notes in computer science, Vol. 6899, pp. 304–327). Berlin: Springer.
Scheider, S., & Stasch, C. (2015, forthcoming). The semantics of sensor observations based on attention. In G. Marchetti, G. Benedetti, & A. Alharbi (Eds.), *Attention and meaning: The attentional basis of meaning*. New York: Nova.
Shirky, C. (2009). Ontology is overrated. http://www.shirky.com/writings/ontology_overrated.html.
Steels, L. (1997). Constructing and sharing perceptual distinctions. In M. Someren & G. Widmer (Eds.), *Machine learning: ECML-97* (Lecture notes in computer science, Vol. 1224, pp. 4–13), Berlin/Heidelberg: Springer.
Steels, L. (2002). Grounding symbols through evolutionary language games. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 211–226). New York: Springer.
Steels, L. (2003). Social language learning. In M. Tokoro & L. Steels (Eds.), *The future of learning* (pp. 133–162). Amsterdam: MIT.
Steels, L. (2008). The symbol grounding problem has been solved. So what's next? In M. de Vega (Ed.), *Symbols and embodiment: Debates on meaning and cognition*, chap 12. Oxford: Oxford University Press.
Stuckenschmidt, H., & van Harmelen, F. (2003). *Information sharing on the semantic web*. Heidelberg: Springer.
Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge: Harvard University Press.
Wittgenstein, L. (2003). *Philosophische Untersuchungen*. Frankfurt a. M.: Suhrkamp.

Chapter 7
Conceptual Spaces and Computing with Words

Janet Aisbett, John T. Rickard, and Greg Gibbon

Abstract The purpose of this paper is to explore synergies and gaps in research in Conceptual Spaces (CS) and Computing with Words (CWW), which both attempt to address aspects of human cognition such as judgement and intuition. Both CS and CWW model concepts in term of collections of properties, and use similarity as a key computational device. We outline formal methods developed in CWW for modelling and manipulating constructs when membership values are imprecise. These could be employed in CS modelling. On the other hand, CS offers a more comprehensive theoretical framework than CWW for the construction of properties and concepts on collections of domains. We describe a specific formalism of CS based on fuzzy sets, and discuss problems with it and with alternative methods for aggregating property memberships into concept membership. To overcome the problems, we present a model in which all constructs are fuzzy sets on a plane, and similarity of two constructs is an inverse function of the average separation between their membership functions.

7.1 Introduction

This paper concerns synergies between Conceptual Space (CS) modelling based on the work of Gärdenfors (2000) and some recent work in fuzzy sets, in particular that known as Computing with Words (CWW). The latter term was coined by Zadeh in 1996 to represent the toolbox of techniques that might be developed for computing when input/output is imprecise words rather than numbers (Mendel et al. 2010; Beliakov et al. 2012; Zadeh 1999, 2012). CWW employs a vocabulary of words or terms, each of which is associated with a fuzzy set on a domain (Mendel 2007a, b). The vocabulary items thus act in the same way as the names or labels of properties or

J. Aisbett (✉) • G. Gibbon
School of Design, Communication and Information Technology, The University of Newcastle, Newcastle, Australia
e-mail: janet.aisbett@gmail.com; greg.gibbon@gmail.com

J.T. Rickard
Distributed Infinity Inc., Larkspur, CO, USA
e-mail: terry.rickard@reagan.com

F. Zenker, P. Gärdenfors (eds.), *Applications of Conceptual Spaces*,
Synthese Library 359, DOI 10.1007/978-3-319-15021-5_7

concepts in CS. The value taken by a membership function can be interpreted as the degree to which a domain element represents the property that the word/term labels.

Both CS and CWW attempt to address aspects of human reasoning such as judgement and intuition. Both use similarity as a key computational device, and both model concepts loosely as a collection of properties. Unlike CS, CWW has a formal method to deal with functions of graded properties, and allows for imprecision in the membership values of domain elements, that is, allows for vague links between domain elements and properties. The next section describes these mechanisms, and, as an example of CWW decision making systems, outlines an architecture called the Perceptual Computer.

The fuzzy sets with which CWW vocabulary words are associated are almost always defined on a scale. CWW does not have the theoretical foundation of CS for modelling properties and concepts as regions on (collections of) domains, and *ad hoc* methods are employed to determine the membership of an entity in a concept given its membership in a collection of properties. The third section of this paper therefore looks at a fuzzy set-based formalism of Conceptual Spaces in which concepts are equipped with structure motivated by Gärdenfors (2000: 105) and which directs how concept membership is computed from property memberships. Unfortunately this structure was found to be practically flawed, and so the section presents alternative approaches to aggregating property memberships which cater for, *inter alia*, properties that are necessary or sufficient in the definition of a concept. The *ad hoc* selection of aggregation methods used in both CWW and CS can be eliminated using modelling presented in our penultimate section. In this, all conceptual constructs have a uniform representation, as fuzzy sets on the same domain, and are compared with the one form of similarity measure.

7.2 Fuzzy Sets, CWW and Perceptual Computing

We will equate fuzzy sets with their membership functions $f : X \to [0, 1]$ from domain X (usually a metric space) into the unit interval, and let $M(X)$ denote the space of fuzzy sets on X. In CWW, domains are typically a scale, such as height or preferences, and the fuzzy set models the referent of a word or a term (although CWW researchers sometimes loosely call the fuzzy set a model for the vocabulary item). These fuzzy models are obtained theoretically or empirically (e.g. Norwich and Turksen 1984; Mendel 1999; Turksen 2002; Beliakov et al. 2012). Membership functions on scales are usually constructed to be mono-peaked and continuous, such as Gaussian or trapezoidal functions.

7.2.1 The Extension Principle

The fundamental computational tool in the CWW toolbox, as in fuzzy sets and systems generally, is the Extension Principle (Zadeh 1975). The Extension Principle

specifies how to extend a multivariate function $g : X_1 \times \cdots \times X_n \to Y$ to a function from the product space of fuzzy sets on the domain components to the space of fuzzy sets on the range. Specifically, fuzzy sets $f_i \in M(X_i)$, $i = 1, \ldots, n$, are mapped under g to the fuzzy set on Y defined by

$$g(f_1, \ldots, f_n)(y) = \sup_{(x_1, \ldots, x_n) \in X_1 \times \cdots \times X_n} \{\min\{f_1(x_1), \ldots, f_n(x_n)\} : g(x_1, \ldots, x_n) = y\};$$

$$g(f_1, \ldots, f_n)(y) = 0, y \notin g(X_1, \ldots, X_n), \quad y \in Y. \tag{7.1}$$

That is, the membership of a point $y \in Y$ is selected to be the supremum of the membership values of elements in the inverse image of y, where the membership value of a vector is the minimum of the membership values of the components.

As a simple example of the application of the Extension Principle, consider a complex graded concept, *edible*, in the context of an apple, and suppose it is defined on two domains, i.e., $edible : X_1 \times X_2 \to [0, 1]$, where, say, the domains are size and hue (greenness). Then (7.1) allows us to specify the degree to which an apple which is *small* and *greenish* is *edible*, as the fuzzy set h on the unit interval given by

$$h(y) = \sup_{x_i \in X_i, i=1,2} \{\min\{small(x_1), greenish(x_2)\} : edible(x_1, x_2) = y\}, \tag{7.2}$$

where we have identified properties with their membership functions. Conversely, the degree to which the concept *edible* coincides with the conjunction of the properties *small* and *greenish* in the context of an apple is the fuzzy set with membership function

$$h'(y) = \sup_{x_i \in X_i} \{edible(x_1, x_2) : \min\{small(x_1), greenish(x_2)\} = y\}. \tag{7.3}$$

Figure 7.1 illustrates x_1-x_2 contours of equi-membership in *edible* (dashed contours) and *small & greenish* (solid contours) which are used in computing (7.2) and (7.3), and indicates $h(y), h'(y)$ in the case in which the contours enclose regions of higher membership value.

7.2.2 Allowing for Imprecision in Membership Functions

Mendel (1999) observed that "[w]ords mean different things to different people" in suggesting that CWW should use type-2 fuzzy set representations which allow for imprecision in word membership functions. A type-2 fuzzy set can be regarded as a function $f : X \times [0, 1] \to [0, 1]$, often written in terms of the so-called secondary membership functions, $f_x : [0, 1] \to [0, 1]$ where $f_x(y) = f(x, y), x \in X, y \in [0, 1]$ (Mendel and John 2002). The main argument against use of type-2 fuzzy

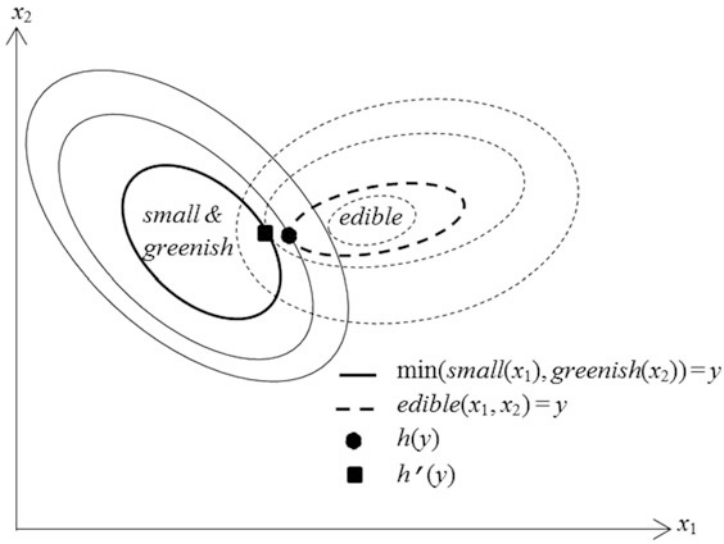

Fig. 7.1 Membership value contours for concepts *edible* (*dashed lines*) and *small & greenish* (*solid lines*). The contours are followed to compute the membership values in Eqs. (7.2) and (7.3) (shown as solid circle and square respectively) (See text for details)

sets concerns whether the considerable complexity of computation is justified by the results (e.g. Greenfield et al. 2009). Many applications adopt an intermediate position, using interval type-2 fuzzy sets in which, for each $x \in X$, a subinterval $\left[\underline{x}, \overline{x}\right]$ of $[0, 1]$ can be found such that $f_x(y) = 1, y \in \left[\underline{x}, \overline{x}\right]$ and $f_x(y) = 0$ otherwise. (An interval type-2 fuzzy set is also known as an interval-valued fuzzy set.)

Empirical interval type-2 fuzzy set representations of abstract and comparative concepts such as *pleasing* or *tall man*, quality judgements such as *very bad* or *good*, and quantities such as *very little* or *moderate amount* have been obtained on scales (Turksen 2002; Wu 2009). These may be elicited by asking multiple participants to nominate an interval on the scale which best describes the meaning of each word, then using an algorithmic procedure (Wu et al. 2012) to estimate the upper and lower bounds of an interval type-2 fuzzy representation of the word. Figure 7.2 illustrates some comparative quantity words on a scale of 0 to 10 (horizontal axis). As is usual in depictions of interval type-2 fuzzy sets, the shaded area depicts $\{(x, y) \in X \times [0, 1] : f_x(y) = 1\}$, which is referred to as the "footprint of uncertainty" (FOU).

Such sets are manipulated in CWW, using (7.1), to answer questions such as "Are Swedes tall?" on the basis of data provided as fuzzy sets (Zadeh 2012). The Extension Principle prescribes how to extend functions defined on fuzzy sets to type-2 fuzzy sets. The trivial example of the union of sets serves to indicate the

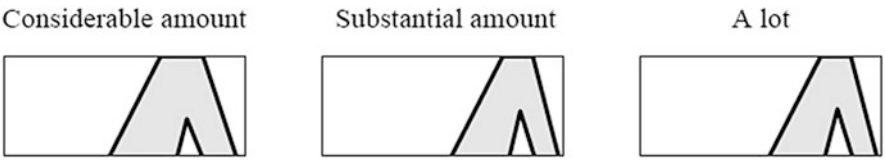

Fig. 7.2 (From Wu and Mendel 2010) Interval type-2 representations of comparative quantity words on a scale of 0 to 10 (horizontal axis). The vertical axis is the primary membership value range of 0 to 1. *Solid lines* show the upper and lower bounding functions on membership intervals

additional complexity of type-2 computations. Given graded properties *tiny* and *small* on a size scale X, say, then $tiny \cup small$ has membership value at $x \in X$ of $\max\{tiny(x), small(x)\}$. If these properties are represented by type-2 fuzzy sets (call them $tiny\text{-}2, small\text{-}2$), then by (7.1) the union is the type-2 fuzzy set with membership function

$$tiny\&small\,(x, y) = \sup_{y', y'' \in [0,1]} \{\min\{tiny\text{-}2\,(x, y'),\ small\text{-}2, (x, y'')\} : \max\{y', y''\} = y\}. \tag{7.4}$$

Thus computing the membership function of a union entails an optimisation problem over the set $\{y', y'' \in [0, 1] : \max\{y', y''\} = y\}$ for each point $x \in X$. In the case of interval type-2 fuzzy sets, however, it is easy to see that the FOU of the union is readily computed.

The next subsection illustrates the use of such operations in a processing architecture designed to manipulate interval type-2 fuzzy sets which model the referents of words.

7.2.3 Computing with Words

The following architecture modifies a conventional interval type-2 fuzzy logic system so as to take in linguistic input and provide linguistic output, for which reason it has been called a Perceptual Computer (Mendel and Wu 2010, 2012; Wu 2009). It could equally be viewed as a CS categorization system for categories described in terms of fuzzy graded properties.

An encoder at the front end of the Perceptual Computer maps vocabulary words into interval type-2 fuzzy representations on numerical domains. A decoder at the back end reverses the process. The vocabulary is represented by interval type-2 fuzzy sets of the form depicted in Fig. 7.2, or by versions of these truncated at the endpoints of the domain. Mendel and colleagues contend these are natural forms for linguistic representation (Mendel and Wu 2010: 1550).

At the heart of the Perceptual Computer is an "engine" which propagates the imprecision inherent in the linguistic input in a mathematically principled manner. The engine must combine the interval type-2 fuzzy sets representing observations about a set of properties so as to validly describe how the observed entity relates to a more complex concept. This may be done through a set of rules, which, as in all such decision systems, encapsulate knowledge of interest and may be built up from training examples.

In this case, rules involve fuzzy predicates and are of the form

$$R_i : \text{ if } Z_1 \text{ is } F_{i\,1} \text{ and } .... \text{ and } Z_k \text{ is } F_{i\,k_i} \text{ then } Y \text{ is } G_i \tag{7.5}$$

where $F_{i\,j}$, G_i, $i = 1, \ldots, n; j = 1, \ldots, k_i$ are interval type-2 fuzzy sets. Presented with a description of an entity in terms of the system vocabulary, the Perceptual Computer encodes the words into the associated interval type-2 fuzzy sets Z_j. The degree to which Z_j is F_{ij} is a measure of the similarity of the input to the antecedent F_{ij}, and is typically computed as a crisp number, while the firing level T_i of the rule R_i is typically taken to be the minimum or the product of the component memberships (Mendel and Wu 2010). Assuming the consequent fuzzy sets G_i in (7.5) are all defined on the same domain, the Perceptual Computer produces an averaged output, usually the weighted average $\sum_i T_i G_i / \sum_i T_i$ where the weights are the firing levels of the rules. The output may then be decoded into the word most similar to it in an output vocabulary, in what Zadeh (2012) calls "linguistic approximation." For example, this word may be a quality judgement about how well the observation accords with a concept.

The key difference between a Perceptual Computer and an interval type-2 fuzzy logic system is, in the opinion of Mendel and Wu (2010: 1560), the fact that the averaged output is of the same form as the input words, and so can be validly compared with the words in the vocabulary.

There are variants of the Perceptual Computer. If all input words are comparable (for example, they describe quality judgements such as *poor*, *good*, etc. about how well an observation satisfies a property), then instead of employing rules the engine may compute weighted averages directly on the word encodings, as $Y_i = \sum_j W_{ij} Z_j / \sum_j W_{ij}$. Here, weights themselves are vocabulary words which may be interpreted as reflecting the importance or, in CS terminology, the salience of the properties or domains, and may be associated with interval type-2 fuzzy models. In any case, the weighted average requires interval type-2 fuzzy arithmetic, which must employ (7.1).

Amongst the applications considered by Mendel and Wu (2010) are support systems for social judgements, such as in a dating advisory system, and for financial decisions. A commercial investment advisory system called Discovery Investing Scoreboard (Rickard et al. 2012) based on the Perceptual Computer employs crowd-sourced rules and more complex similarity measures.

7.3 Fuzzy Conceptual Spaces and the Role of Aggregation

The previous section has described CWW work in which (type-2) fuzzy properties are combined via conjunction, and other work in which salience is modelled through words which reflect the importance of a property. However, CWW does not have a disciplined way of building complex conceptual structures.

Gärdenfors (2000: 105) suggests that a natural concept is represented as a set of regions from different domains together with their salience and correlations. This motivated Rickard (2006) to define a concept on a set of n properties to be an $n \times n$ matrix with entries in $[0, 1]$. Diagonal elements are the properties' salience and off-diagonal entries are the co-occurrence of the pair of properties in the concept. In order to compare an observation with a concept (e.g. to classify it), the observation is converted to a pseudo-concept in which the co-occurrence of two properties is deemed to be the smaller of the observation's property membership values. This work was extended in CS modelling using fuzzy sets (Rickard et al. 2007) and type-2 fuzzy sets (Rickard et al. 2010; Aisbett and Rickard 2013) which we now briefly describe.

7.3.1 CS with Fuzzy Constructs

Take a (fuzzy) property to be a (type-2) fuzzy set on a domain together with a name or label, and a context I to be a (fuzzy) set of properties. Observations are either values or fuzzy sets on a collection of domains. When a property is a type-2 fuzzy set with membership function $g : X \times [0, 1] \to [0, 1]$ and an observation on that domain is a fuzzy set $f : X \to [0, 1]$, then the membership of the observation in the property is a fuzzy set on the unit interval, $g(f)(y) = \sup_{x \in X} \{\min \{g(x, y), f(x)\}\}$.

Suppose a context I contains n properties. Then a concept in that context is a (type-2) fuzzy set on $I \times I$ in which each membership value denotes the co-occurrence of the pair of properties in the concept, or, for diagonal elements, their salience. Rickard and colleagues define similarity using directional subsethood, which measures the relative overlap of fuzzy sets and does not require a metric on the domains. Specifically, the subsethood of a fuzzy set G in a fuzzy set H is the ratio of the average membership of the fuzzy intersection $G \cap H$ to the average membership of G over the domain. Subsethood of type-2 fuzzy sets is derived from this using the Extension Principle. While there is no analytical form for subsethood of type-2 fuzzy sets, computationally efficient algorithms have been developed for the interval type-2 case (e.g. Nguyen and Kreinovich 2008).

Like any similarity measure, subsethood enables comparison of concepts defined in the same context, and of properties defined on the same domain. Hence concepts defined on different but related contexts can also be estimated using the property overlap.

The constructs were tested in classification tasks. For these, the property co-occurrences which define concepts (classes) are computed as their directional subsethood over the set of training observations. Test observations are converted to pseudo-concepts and compared with the candidate concepts using subsethood, and the subsethoods are ranked. On traditional machine learning datasets such as the "Mushroom" and "Cleveland Heart", the fuzzy CS method performed at least as well as nearest neighbour and SVM (Laeta 2007). More realistic applications investigated with the type-2 modelling included land use suitability and change assessments and prediction of share price behaviour (Laeta 2008; Aisbett and Rickard 2013).

While it can be argued that subsethood is a natural way of comparing fuzzy sets, in practice there are computational problems (even before going to type-2) because subsethood involves a ratio in which the denominator can become small. In addition, salience is conveyed through only n of the n^2 terms used in computing similarity (whether through subsethood or any other conventional measure on fuzzy sets defined on $I \times I$) and can be "swamped" by the co-occurrences.

Other aggregation approaches therefore are recommended to define a concept in terms of a set of (fuzzy) properties, and to model the relationship between an observed entity and a concept when observations are a set of property membership values or fuzzy sets.

7.3.2 Dealing with Complex Structures

Computing an observation's similarity to (equivalently, membership in) a complex structure by aggregating similarity to (membership in) a collection of properties is problematic. For example, the weighted average is clearly unsatisfactory in aggregating property similarities/memberships for structures in which one property is necessary or is sufficient. Beliakov observes that the initial interpretation of the inputs is lost in the complexity of an aggregation process (Beliakov et al. 2012). And, as has often been pointed out, a "common scale" assumption is behind even the simplest aggregation, and if weightings do not achieve this then aggregation cannot produce valid memberships for composite concepts (Bellman and Giertz 1973; Lee 2003).

Nevertheless, a wealth of aggregation methods has been employed in diverse applications relevant to CWW and to CS (e.g. Yager 1993; Ralescu and Ralescu 1997; Beliakov et al. 2007). The fuzzy weighted power means (WPMs) are one such family, and generalise the weighted average incorporated into some versions of the Perceptual Computer. Given a set of properties, suppose the membership of an observation in the ith property is v_i. Given a concept C suppose the salience of the ith property is w_i. Then for any power $p, -\infty \leq p \leq \infty$, the WPM of the membership is

$$L(\mathbf{v}, \mathbf{w}; p) = \left(\sum_{i=1,n} w_i v_i^p \Big/ \sum_{i=1,n} w_i \right)^{1/p}, \mathbf{v} = (v_1, \ldots, v_n), \mathbf{w} = (w_1, \ldots, w_n). \quad (7.6)$$

To see that the WPM is a plausible candidate for the membership of the observation in the concept, note that as the power varies from minus to plus infinity, the aggregate varies from the minimum to the maximum of the property membership values (Dujmović 2007). Thus the concept tends towards the conjunction of the properties as $p \to -\infty$ and tends towards disjunction as $p \to \infty$. Moreover, when $p < 0$ the aggregate is zero whenever any property membership is zero, so that each property is mandatory, in an obvious sense. The choice of the power is thus a key part of the modelling of the relationship of the concept to the properties.

A more powerful mechanism is provided by introducing complementary aggregations of the form $1 - L(\mathbf{1} - \mathbf{v}, \mathbf{w}; p) = 1 - \left(\sum_{i=1,n} w_i (1 - v_i)^p \Big/ \sum_{i=1,n} w_i \right)^{1/p}$ (Dujmović and Larsen 2007). When $p < 0$, this complement takes value 1 when any of the property memberships are 1, so that each property is sufficient. Subsets of the properties can be grouped according to their role in the definition of the concept, and assigned different powers in the aggregation, as for example in

$$0.5 \left(\left(\sum_{i \in I_1} w_i v_i^p \Big/ \sum_{i \in I_1} w_i \right)^{1/p} + 1 - \left(\sum_{i \in I_2} w_i (1 - v_i)^q \Big/ \sum_{i \in I_2} w_i \right)^{1/q} \right), \quad (7.7)$$

$I_1 \cup I_2 = \{1, \ldots, n\}$, $p \neq q$. Fuzzy and type-2 fuzzy versions of this have been applied to CWW decision support systems (Rickard et al. 2011).

Another class of aggregation functions of potential interest in computing concept membership from a set of property memberships is the Tsallis q-exponential (Tsallis 2009). Rickard and Aisbett (2014) claim this aggregation can facilitate modelling of threshold-dependent necessary or sufficient aggregations of properties in concepts, which they call "all-out" or "all-in" reasoning. These authors illustrated the flexibility of employing q-exponential-based aggregation in CWW on the investment judgment advisor system presented by Mendel and Wu.

While we believe that a range of aggregation models is necessary in both CWW and CS to match characteristics of a complex concept, we know of no comprehensive theoretical or practical framework for the choice of model.

7.4 A Uniform Conceptual Space

Applications described so far have been targeted at computer-based advisory systems. CS modelling can also explore understanding of human cognition (Gärdenfors 2000). To this end, Aisbett and Gibbon (2005) developed a CS process model along a "cortical map" metaphor. The treatment addresses issues of *ad hoc* definition of aggregation, and the common scale problem for membership values.

7.4.1 The "Cortical Map" Model

All domains in the "cortical map" model are subsets of a finite connected region X of the real plane, and include a distinguished region used for indexing lexical data. Concepts, properties and observations are modelled as fuzzy sets (i.e., as elements in $M(X)$; Aisbett and Gibbon (2005) called them images) and context is defined as a subregion of X, i.e., as a crisp member of $M(X)$. The specification of X as a planar region is motivated by the notion of cortical layers, but is mathematically justified by the fact that the lowest dimensioned region which cannot be disconnected by the removal of a single point is planar.

Long term memory and a memorization process are modelled, permitting an exemplar-based definition of a property or a concept (properties and concepts are not distinguished in the "cortical map" model since a property can always be subdivided). Specifically, a property/concept is an index or lexical label plus a collection of fuzzy sets stored in memory which are similar to the concept label on the index domain. Although a fuzzy set prototype representing the property/concept is formed dynamically, the exemplar-base definition is fundamentally different to a prototype-based system because the transient prototype is dependent on context, as well as on the current contents of memory. For details, see Aisbett and Gibbon (2005).

Because X is a region on the real plane, a metric based measure of similarity can be employed, assuming the membership functions in $M(X)$ are integrable. Given a context $D \subseteq X$, $M(D)$ (the set of fuzzy sets on D) is a metric space under the Hamming distance $d_D(f,g) = \int_D |f(x) - g(x)|\, dx$. Similarity in the context D is

$$s_D(f,g) = \exp(-c d_D(f,g)) = \exp\left(-c \int_D |f(x) - g(x)|\, dx\right), \tag{7.8}$$

where c is a scaling factor. Since distance over the union of disjoint domains is obviously additive, the similarity of the union over mutually disjoint domains D_j, $j = 1, \ldots, l$, is thus the product of similarities $s_{\cup D_j}(f,g) = \prod_{j=1,l} s_{D_j}(f,g)$. Unlike the externally determined context used in the systems described in previous sections, context is determined by an attention setting process which employs a working memory consisting of copies of $M(X)$. Details are in Aisbett and Gibbon (2005).

Any modelling must link to observable external domains such as hue, size or taste. As discussed earlier, CWW researchers have obtained (interval type-2) fuzzy set models on scalar domains with membership functions constrained to belong to a given family. However, that reproducible representations may be obtained on more complex and integral domains is suggested by experiments reported in Aisbett and Gibbon (2003). Figure 7.3b is a fuzzy set representation of the plasticine cube pictured in (a), where the outlined square depicts a region D of X and the fuzzy set value on D is depicted by greyscale where 1 = black. The physical domains, as well as the family of fuzzy sets that could be used to represent values on these domains, were constrained in experiments designed to test, amongst other things,

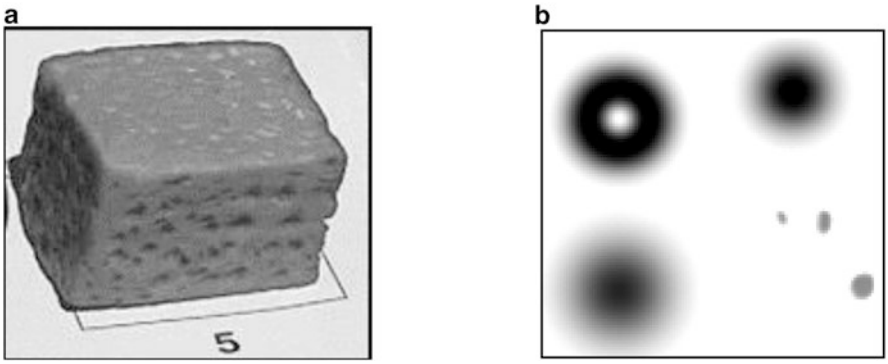

Fig. 7.3 (Aisbett and Gibbon 2003) (**a**) Real world object which can be described in many ways. (**b**) Representation of the object as a fuzzy set on a plane. The quadrants represent (clockwise from *top left*) size & weight, hue, texture & feel, and shade of the object. *Dark* indicates high membership value

how such fuzzy set representations of an object were correlated between participants (Aisbett and Gibbon 2003). Perceptual similarity was found to be well modelled by the negative exponential of the Hamming distance on the constrained families of fuzzy sets. This finding was essential, as mappings $\phi_E : E \to M(X)$ from external domains E should preserve order.

When E is a scale with a lower bound, Aisbett and Gibbon (2005) assumed mappings ϕ_E satisfy

$$a \leq b \Rightarrow \phi_E(a)(x) \leq \phi_E(b)(x), a, b \in E, x \in X. \tag{7.9}$$

For example, ϕ_E might map a finite interval into a crisp set of concentric discs $D_{o,r}$ of radius $r \in [0, 1]$ centred at $o \in X$ so that the radius is a monotonic function of the interval values. A fuzzy set on the external domain maps to a fuzzy set on $M(X)$ using the Extension Principle, which by (7.1) has zero membership except on the image of ϕ_E. Standard centroid defuzzification reduces this fuzzy set to an element of $M(X)$. In the case that $\phi_E(a) = D_{o,a}$ then this element is the fuzzy set on X with membership function

$$\begin{aligned} \mu(x) &= \int_E \mu_A(r)\phi_E(r)(x)dr \Big/ \int_E \mu_A(r)dr \\ &= \int_{r \geq s} \mu_A(r)dr \Big/ \int_E \mu_A(r)dr, \text{ where } |x - o| = s. \end{aligned} \tag{7.10}$$

Figure 7.4a depicts a trapezoidal membership function and the membership function of its mapping along any radial from the centre $o \in X$. Figure 7.4b depicts two elements of $M(X)$, each of which is composed of the defuzzified mapping of two external domains on which membership functions were truncated Gaussians.

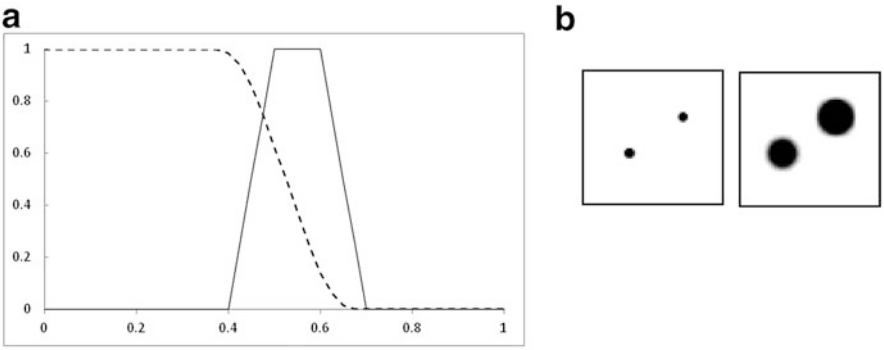

Fig. 7.4 (**a**) *Solid line* depicts a fuzzy membership function on a scale (horizontal axis) representing a property such as *a considerable amount*. The mapping of this fuzzy set into $M(X)$ has membership value shown by the *dotted line* as a function of radial distance (horizontal axis) from centre point $o \in X$. (**b**) The representation of two observations, each taking values in the same two external domains

The Gaussians associated with the left image (membership function) had smaller centroids and smaller means than those associated with the right image.

7.4.2 Categorization Using the "Cortical Map" Representation and the Generalized Context Model

The process model of the "cortical map" system performs differentiation, categorisation, aggregation and composition. Refer to Aisbett and Gibbon (2005) for details of the management of working memory, including the exponential decay of activation on a layer and its augmentation by memories activated by the current observation.

Here, we simply want to outline the model's treatment of categorization after each of m candidate concept names has been associated with an exemplar set of training memories K_i, $i = 1, \ldots, m$. In the test phase, a "prototypical" fuzzy set is accumulated which is the average of all the memories in K_i. According to the process model, presentation of a test observation $f_0 : X \to [0, 1]$ at time t sets attention according to the absolute difference between the observation and the mean of the "prototypes"; that is, attention is the fuzzy set on X with $a_0(x) = \left| f_0(x) - \sum_{i=1,m} f_i(x)/m \right|$. Thresholding thus produces a time-varying region in attention, call it $A_0 = \left\{ x : \left| f_0(x) - \sum_{i=1,m} f_i(x)/m \right| \geq \tau \right\}$, which we assume to have positive area.

Memories are activated according to their similarity to f_0 in A_0. At a time T after the presentation of the observation, total activation is the fuzzy set whose membership at x is

$$\exp(-dT)\, f_0(x) + \lambda(T) \sum\nolimits_{g \in \text{Mem}} s_{A_0(t)}(f_0, g)\, g(x) \tag{7.11}$$

where Mem is the set of memories, d is a decay constant and $\lambda(T)$ is a normalising constant that essentially conserves overall activation. Unlike the observation f_0 which has no concept label attached, the activated memories g carry lexical data, and the response is assumed to be the concept i whose index region L_i is most activated. Using (7.11) and (7.8) this concept is seen to be that which maximises

$$\sum\nolimits_{g \in K_i} s_{A_0}(f_0, g) \int_{L_i} g(x)dx = \sum\nolimits_{g \in K_i} \exp\left(-c \int_{A_0} |g(x) - f_0(x)|\, dx\right) \times \int_{L_i} g(x)dx. \tag{7.12}$$

It is instructive to relate this process to the classic exemplar-based Generalized Context Model (GCM) of Nosofsky (1986). Suppose m concepts are determined by values in a set of k psychological dimensions, and that $K_i = \{g \equiv (g(1), \ldots g(k))\}$ is the set of memorised exemplars of concept i. Denote the value of an observation (also called a stimulus or probe) on dimension j by $f_0(j)$. Then Nosofsky (1986: 44), in the case of a city block metric, assigns the observation to the concept i according to

$$h_i = b_i \sum\nolimits_{g \in K_i} \exp\left(-c \sum\nolimits_j w_j\, |g(j) - f_0(j)|\right) \tag{7.13}$$

where b_i is a bias parameter and w_j are salience weights.

The GCM can be represented by the "cortical map" modelling because the right hand term in (7.12) can be designed to be proportional to (7.13). To see this, first note that, because bias is assumed to be preferential memorisation of some concepts, it will affect the activation of the concept labels. Thus, for each memory g in K_i we can reasonably set

$$\int_{L_i} g(x)dx = b_i. \tag{7.14}$$

Next, note that perceived stimuli on any dimension j must lie in a bounded interval, call it $\left[0, \overline{b}_j\right]$. Hence for some $s > 0$ we can choose points $o_j, j = 1, \ldots, k$ in A_0 and disjoint disks $D_{o_j, s\sqrt{w_j \overline{b}_j}} \subset A_0$ centred at o_j of radius

$s\sqrt{w_j \overline{b}_j}$. Define $\varphi : \left[0, \overline{b}_1\right] \times \cdots \times \left[0, \overline{b}_k\right] \to A_0$ to take an observation or memory $(a_1, \ldots, a_k)$ to the crisp set $\cup_{j=1,k} D_{o_j, s\sqrt{w_j a_j}} \subseteq A_0$. Then

$$\int_{A_0} \left|\phi\left(a_1, \ldots, a_k\right) - \phi\left(a'_1, \ldots, a'_k\right)\right| dx = \pi s^2 \times \sum\nolimits_{j=1,k} w_j \left|a_j - a'_j\right|, \quad a_j, a'_j \in \left[0, \overline{b}_j\right]. \tag{7.15}$$

Substituting (7.14) and (7.15) into the right hand term in (7.12) produces the term on the right side of (7.13) modulo the factor πs^2.

Using the fuzzy set representations and process model of the "cortical map" system, we re-created simulations originally performed by Nosofsky (1986) to explain experimental categorisation results, and obtained comparable results. Figure 7.4b depicts representations of two perceived stimuli as used in our simulations; each carries information about values on the two experimental dimensions.

In summary, the notable characteristic of the "cortical map" model presented in this section is its uniform representation of dimensions and domains (as subsets of X) and of properties, concepts and observations (as subsets of $M(X)$). Aggregation is simply summation. At the same time, the modelling provides a more fluid notion of these CS constructs, since a domain can be any part of X that happens to be in the region in attention at a certain time, and a property/concept is created from memories according to attention.

The model has few parameters, with the scale parameter c the only design choice as to aggregation. Imprecision and uncertainty are modelled identically across all properties/concepts, resolving common scale issues. However, substantial modelling effort may be required to obtain valid mappings from external domains such as texture into $M(X)$ as well as to obtain useful representations of complex concepts such as edible food. Situations in which such effort might be repaid need to be clarified.

7.5 Conclusion

Meso-level representation like CS involves associating words or symbols with structures which can be manipulated by computers. Gärdenfors argued that these structures should be geometrical in nature. Fuzzy sets, as functions into the unit interval (or in the case of type-2 fuzzy sets, into the space of fuzzy sets on the unit interval) come equipped with geometrical structure by virtue of the order and distance on the unit interval. CWW is a program to develop tools to "compute" with (type-2) fuzzy set models of words/terms. This paper has aimed to show that CWW and CS research has many potential crossovers.

While it is straightforward to define structures of CS in terms of (type-2) fuzzy sets, it remains unclear how to best model the relationship between a concept

and the properties used to describe it. We found practical problems with an interpretation of Gärdenfors' description of a concept (in a context) as a set of properties, their saliences and their co-occurrences. Investigating alternative ways to relate property membership to concept membership opens up the huge literature on aggregation, amongst which can be found flexible methods with appealing characteristics. However, there is no current framework to guide selection of the appropriate aggregation structure for a particular concept, leaving open questions about the validity of the modelling.

A different approach is offered by a representation in which all the constructs of CS are defined uniformly, as fuzzy subsets of the plane. Dimensions and domains are crisp sets (regions) defined in response to attention. Properties/concepts are formed on-the-fly and modelled as fuzzy regions, where membership value reflects imprecision or uncertainty. Similarity is modelled as the Hamming distance between fuzzy sets, and is the heart of a uniform processing model. However, this approach puts modelling effort into the initial mapping of external domain values onto fuzzy sets on the plane, and it is not clear under what circumstances external domains can be coherently represented. Another open question is how to generate representations of a complex concept, whether through training examples, empirical methods and/or theory.

References

Aisbett, J., & Gibbon, G. (2003). *Preserving similarity in representation: A scheme based on images*. In Proceedings of the joint international conference on cognitive science, Sydney.

Aisbett, J., & Gibbon, G. (2005). A cognitive model in which representations are images. *Cognitive Systems Research, 6*(4), 333–363.

Aisbett, J., & Rickard, J. T. (2013). Type-2 fuzzy sets and conceptual spaces. Advances in type-2 fuzzy sets and systems. *Studies in Fuzziness and Soft Computing, 301*, 113–129.

Beliakov, G., Pradera, A., & Calvo, T. (2007). *Aggregation functions: A guide for practitioners* (Studies in fuzziness and soft computing 221, pp. 297–304). Berlin: Springer.

Beliakov, G., Bouchon-Meunier, B., Kacprzyk, J., Kovalerchuk, B., Kreinovich, V., & Mendel, J. M. (2012). Computing With Words (CWW): Role of fuzzy, probability and measurement concepts, and operations. *Mathware & Soft Computing Magazine, 19*(2), 27–45.

Bellman, R. E., & Giertz, M. (1973). On the analytic formalism of the theory of fuzzy sets. *Information Sciences, 5*, 149–156.

Dujmović, J. (2007). Continuous preference logic for system evaluation. *IEEE Transactions on Fuzzy Systems, 15*(6), 1082–1099.

Dujmović, J., & Larsen, H. L. (2007). Generalized conjunction/disjunction. *International Journal of Approximate Reasoning, 46*, 423–446.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press.

Greenfield, S., Chiclana, F., Coupland, S., & John, R. (2009). The collapsing method of defuzzification for discretised interval type-2 fuzzy sets. *Information Sciences, 179*(13), 2055–2069.

Laeta P/L. (2007). *Alternative conceptual space constructs in relation to computing with words -1*. Report to Lockheed Martin – Integrated Systems & Solutions Littleton CO.

Laeta P/L. (2008). *Alternative conceptual space constructs in relation to computing with words-2*. Report to Lockheed Martin – Integrated Systems & Solutions Littleton CO.

Lee, J. W. T. (2003). Ordinal decomposability and fuzzy connectives. *Fuzzy Sets and Systems, 136*(2), 237–249.

Mendel, J. M. (1999). Computing with words, when words can mean different things to different people. In *Proceedings of 3rd international ICSC symposium on fuzzy logic and applications* (pp. 158–164), Rochester, NY: ICSC Academic Press.

Mendel, J. M. (2007a). Computing with words: Zadeh, Turing, Popper and Ockam. *IEEE Computational Intelligence Magazine, 2*, 10–17.

Mendel, J. M. (2007b). Computing with words and its relationships with fuzzistics. *Information Sciences, 177*, 985–1006.

Mendel, J. M., & John, R. I. (2002). Type-2 fuzzy sets made simple. *IEEE Transactions on Fuzzy Systems, 10*(2), 117–127.

Mendel, J. M., & Wu, D. (2010). *Perceptual computing*. Hoboken: Wiley.

Mendel, J. M., & Wu, D. (2012). Challenges for perceptual computer applications and how they were overcome. *IEEE Computational Intelligence Magazine, 7*(3), 36–47.

Mendel, J. M., Zadeh, L. A., & Trillas, E., et al. (2010, February). What computing with words means to me. *IEEE Computational Intelligence Magazine, 5*(1), 20–26.

Nguyen, H., & Kreinovich, V. (2008). Computing degrees of subsethood and similarity for interval-valued fuzzy sets: Fast algorithms. In *Proceedings of 9th international conference on intelligent technologies* (pp. 47–55), Assumption University of Thailand.

Norwich, A. M., & Turksen, I. B. (1984). A model for the measurement of membership and the consequences of its empirical implementation. *Fuzzy Sets and Systems, 12*, 1–25.

Nosofsky, R. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39–57.

Ralescu, A. L., & Ralescu, D. A. (1997). Extensions of fuzzy aggregation. *Fuzzy Sets and Systems, 86*, 321–330.

Rickard, J. T. (2006). A concept geometry for conceptual spaces. *Journal of Fuzzy Optimization and Decision Making, 5*(4), 311–329.

Rickard, J. T., & Aisbett, J. (2014). New classes of threshold aggregation functions based upon the Tsallis q-exponential with applications to perceptual computing. *IEEE Transactions on Fuzzy Systems, 22*(3), 672–684.

Rickard, J. T., Aisbett, J., & Gibbon, G. (2007). Reformulation of the theory of conceptual spaces. *Information Sciences, 177*, 4539–4565.

Rickard, J. T., Aisbett, J., Yager, R., & Gibbon, G. (2010). Type-2 fuzzy conceptual spaces. In *Proceedings of IEEE international conference on fuzzy systems* (pp. 1–8), Barcelona: IEEE Inc.

Rickard, J. T., Aisbett, J., Yager, R., & Gibbon, G. (2011). *Fuzzy weighted power means in evaluation decisions*. In Proceedings of world symposium on soft computing, Paper #100, San Francisco, CA.

Rickard, J. T., Berry, M. A., Rickard, T., Morgenthaler, D. A., Berry, C., & Holland, R. (2012). Computing with words for discovery investing: A case study of computational intelligence for the digital economy. In *Proceedings of IEEE international conference on fuzzy systems* (pp. 1–8), Brisbane: IEEE Inc.

Tsallis, C. (2009). *Introduction to nonextensive statistical mechanics*. New York: Springer.

Turksen, I. B. (2002). Type 2 representation and reasoning for CWW. *Fuzzy Sets and Systems, 127*, 17–36.

Wu, D. (2009). *Intelligent systems for decision support*. PhD thesis, University of Southern California.

Wu, D., & Mendel, J. M. (2010). Social judgment advisor: An application of the perceptual computer. In *Proceedings of IEEE international conference on fuzzy systems* (pp. 1–8), Barcelona: IEEE Inc.

Wu, D., Mendel, J. M., & Coupland, S. (2012). Enhanced interval approach for encoding words into interval type-2 fuzzy sets and its convergence analysis. *IEEE Transactions on Fuzzy Systems, 20*(3), 499–513.

Yager, R. R. (1993). A general approach to criteria aggregation using fuzzy measures. *International Journal of Man-Machine Studies, 38*, 187–213.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning – I. *Information Sciences, 8*(3), 199–249.

Zadeh, L. A. (1996). Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems, 4*, 103–111.

Zadeh, L. A. (1999). From computing with numbers to computing with words – From manipulation of measurements to manipulation of perceptions. *IEEE Transactions on Circuits and Systems, 45*, 105–119.

Zadeh, L. A. (2012). *Computing with words: Principal concepts and ideas*. Heidelberg: Springer.

Chapter 8
Self-Organisation of Conceptual Spaces from Quality Dimensions

Paul Vogt

Abstract This chapter presents a discussion on how conceptual spaces can evolve from a set of quality dimensions, and how these spaces can become shared among a population of cognitive agents. An agent-based simulation of Steels' Talking Heads experiment is presented in which virtual agents construct novel concepts, as well as a shared, simplified language from scratch. Simulations demonstrate that the structure of a conceptual space (i.e. from what quality dimensions it is composed) can evolve in a population of communicating agents. It is argued that the underlying mechanisms involve the following factors: the environment of the agents, their embodiment and cognitive capacities, self-organisation, and cultural transmission.

8.1 Introduction

Conceptual spaces are constructed from quality dimensions (Gärdenfors 2000), but how are quality dimensions selected to constitute a particular conceptual space? Is it the result of biological evolution? Or do the conceptual spaces emerge through ontogenetic development? And, if the latter, are they culturally determined and/or constrained through cognition, embodiment, or the ecological niche? I will argue that it is probably a combination of all these factors.

To answer these questions, let me start by briefly recapturing what quality dimensions are and how they constitute conceptual spaces. According to Gärdenfors (2000, p. 6) "the primary function of the quality dimensions is to represent various 'qualities' of objects ... [and] correspond to the different ways stimuli are judged to be similar or different". In visual perception, for instance, these qualities could be feature detectors such as hue, saturation and brightness to represent the conceptual space of colour, edge detectors that may combine to represent a shape, spatial detectors that combine to represent spatial locations, etc. It is beyond doubt that many (if not all) of these quality feature detectors are innate and have evolved biologically. It is also arguable that evolution has selected for

P. Vogt (✉)
Tilburg centre for Cognition and Communication, Tilburg University,
P.O. Box 90153, 5000 LE Tilburg, The Netherlands
e-mail: p.a.vogt@uvt.nl

F. Zenker, P. Gärdenfors (eds.), *Applications of Conceptual Spaces*,
Synthese Library 359, DOI 10.1007/978-3-319-15021-5_8

particular configurations of feature detectors (or quality dimensions) that together form a particular conceptual space, such as the colour space. However, this does not necessarily hold for all possible conceptual spaces.

Take, for example, the space of spatial concepts. Languages have different ways of communicating spatial relations, based on three different frames of reference: relative to the target object (e.g. the box on the left of the tree), intrinsic to the target (e.g. I am in front of the box) or absolute (the box to the north). Often languages have a combination of two or three of these frames of reference, while other languages have only one of these (Majid et al. 2004). There is abundant evidence that the way people categorise and the way they communicate about spatial relations are tightly linked, so the people who speak using a particular frame of reference, also categorise the world that way (Majid et al. 2004; Haun et al. 2011). Such a Whorfian account (Whorf 1956) not only holds for spatial concepts, but also for many other types of concepts (Bowerman and Levinson 2001). Since the speakers of different languages categorise spatial relations so radically different, it is conceivable that their concepts are represented in conceptual spaces made up of different quality dimensions.

The spatial concepts that people use in their language community is clearly learnt, flexible, and depends to some extent on the physical environment (Haun et al. 2011). Tzeltal speakers, for instance, live in a hilly environment and communicate spatial relations in terms of being uphill or downhill. Now, imagine a Tzeltal speaker moving to the Netherlands where there are no hills. At first, he will have difficulty categorising the world in terms of spatial concepts. Nevertheless, he will be able to distinguish that objects are in different spatial locations. If he learns Dutch, he will learn that spatial relations are communicated in a relative frame of reference using left, right, front, etc. Its concepts would be represented in a different conceptual space constructed from a different set of quality dimensions. If he gets a child in the Netherlands and, as long as they will not leave the Netherlands, this child will learn to categorise the world the way Dutch people do in terms of left, right, etc., as well as the concepts of North, East, etc., but although the child may learn to speak Tzeltal, he will not be able to form the concepts of uphill and downhill. (Imagine Dutch speakers talking about things being uphill or downhill, while there are virtually no hills in the Netherlands.) To really being able to form such concepts, the child will need exposure to a hilly environment.

In this chapter, I will demonstrate how a group of virtual (i.e., simulated) robots can acquire various conceptual spaces from a given set of quality dimensions by developing a set of linguistic conventions from scratch through cultural evolution (I will take the biological evolution of conceptual spaces for granted). Before doing that, I will set the theoretical framework of these simulations in which I argue that the following factors are the driving force behind such a development: the environment (or ecological niche), embodiment (i.e. the physical properties of the agent), cognition (e.g., the way concepts are learnt and represented), self-organisation and cultural transmission. The robotic agents are necessarily abstracted away from human agents, so that the results are not directly generalisable to human cognition. The purpose of this chapter is therefore not to present a biologically

plausible model of the formation of conceptual spaces from the basic primitives of quality dimensions, but to illustrate a number of likely mechanisms and properties that could explain how such a development could work.

8.2 The Evolution of Conceptual Spaces

The theoretical framework is developed from an evolutionary linguistics point of view, because of the tight link between linguistic and conceptual structures (Majid et al. 2004; Bowerman and Levinson 2001). In particular, the framework will be based on a hypothesised evolutionary transition from holistic protolanguages to more modern compositional languages (Wray 1998). In order to do that, it is instrumental to define such languages:

Holistic languages are languages in which parts of expressions have no functional relation to any parts of their meanings. For instance, there is no part of the expression "bought the farm" that relates to any part of its meaning "died".

Compositional languages are languages in which parts of expressions do have a functional relation to parts of their meanings and the way they are combined. For instance, the part "John" in "John loves Mary" refers to a guy named John, likewise "loves" and "Mary" have their own distinctive meanings. In addition, this sentence has a different meaning in English when the word-order changes, as in "Mary loves John".

Based on these definitions, it is possible to conceive that if a particular meaning is associated with a holistic utterance, then this meaning could be represented in some N-dimensional conceptual space. However, when the same meaning would be associated with a compositional utterance, then parts of the utterance would be associated with individual concepts c_i, each represented within an n_i-dimensional conceptual space with $n_i \leq N$.

Alison Wray (1998) has argued that protolanguages were essentially holistic in nature and that from these initial stages language has gradually evolved into compositional languages. Although it has been argued that protolanguages were not holistic, but synthetic and instead consisted of multi-word utterances without a particular syntactic structure (Bickerton 1984; Jackendoff 2002), let us assume that Wray is correct. (Without justification, I believe that many of the underlying principles presented in this chapter would hold either way.) Then one could ask the question: what evolutionary mechanism(s) caused this transition? The nativist account would be that the population of language users have adapted biologically to learn and produce compositional languages (Pinker and Bloom 1990). If this occurred through natural selection, this would require that individuals with a particular genetic mutation started using compositional language (at least to some extent), which made them evolutionary more advantageous, thus improving their chances of passing on this mutation, thus increasing the population of individuals using compositional language, etc. Although not impossible, biological evolution

is a rather slow process that would take quite a number of generations before a mutation is spread among the entire population.

An alternative explanation takes the view that cultural evolution was the driving force behind the transition from holistic protolanguage to compositional language. In this viewpoint, put forward by Wray herself and soon adopted by Simon Kirby and colleagues (Brighton and Kirby 2001; Kirby 2001; Kirby et al. 2004, 2008; Kirby and Hurford 2002), the population of language users does not adapt to learn and use compositional languages, but the language adapts itself such that it can be learnt and produced by its users. This is an appealing explanation, not only because language change spreads faster across a population through cultural evolution, but also because when a genetic mutation yields a change in the language that other language users cannot deal with, the mutant language user does not conform to the other users, thus hampering effective communication.

The potential of this cultural evolutionary explanation for this transition has been demonstrated over and over again in computer simulations (Brighton and Kirby 2001; Kirby 2001; Kirby et al. 2004; Kirby and Hurford 2002; Vogt 2005a) and in psycholinguistic experiments (Garrod et al. 2010; Kalish et al. 2007; Kirby et al. 2008). The typical approach in these simulations and experiments is based on iterated learning in which the language of one individual is passed on to a learner from a next generation, who in turn passes on the language to the next generation, and so forth. This thus creates a chain of generations of language users who each acquire the language from the previous generation.

The learners in this model are endowed with a learning mechanism that enables them to discover regular patterns in the input (both in speech and semantics) and when a regularity is discovered, a compositional representation can be constructed and used. This is especially useful when a language user wants to communicate a previously unseen meaning that is composed of several concepts for which the user knows words or utterances to express parts of the meaning, but not the whole meaning holistically. Kirby and colleagues have demonstrated that a transition from holistic languages to compositional language occurs when the language is transmitted through a bottleneck where the next generation needs to communicate about previously unseen meanings. The primary reason for this is that a bottleneck makes the transmission of holistic languages unstable, but not for compositional languages, as illustrated in Fig. 8.1.

The abstractions and assumptions made in the iterated learning model, especially in its computational implementations, however, make it hard to generalise the results. For instance, it is typically assumed that each generation has only one individual and that only the individual from the older generation passes on language to the next generation, thus it rests entirely on vertical transmission. Consequently, the researcher has to impose the transmission bottleneck explicitly. In addition, in most computer simulations the semantics are predefined by the researchers, who thus ensures that there are clear decomposable semantic structures.

A more realistic model would assume a population containing many individuals from different generations, who can each pass on parts of the language to other individuals more akin to oblique and cultural transmission. This is important,

Type	$G(n)$	Utterance	$G(n+1)$
Holistic	toma-[redsquare] tula-[greentriangle] bulo-[greensquare] rino-[redtriangle]	toma-[redsquare] tula-[greentriangle] bulo-[greensquare]	toma-[redsquare] tula-[greentriangle] bulo-[greensquare] ??-[redtriangle]
Compositional	toma-[redsquare] bulo-[greentriangle] buma-[greensquare] tolo-[redtriangle]	to*ma*-[red*square*] **bu**lo-[green**triangle**] **bu***ma*-[**green** *square*]	toma-[redsquare] bulo-[greentriangle] buma-[greensquare] tolo-[redtriangle]

Fig. 8.1 This figure illustrates why holistic languages (upper part) are unstable when a population of generation $G(n+1)$ only observes three of the four utterances from generation $G(n)$'s language (i.e. word-meaning mappings). In this case, if generation $G(n+1)$ wishes to communicate about meaning [redtriangle], then this generation will have to create a new word. If the language were compositionally structured as in the bottom part of this figure, observing the aligning patterns from only three out of four utterances would allow the next generation to reconstruct the entire previous language. Hence transmitting a compositional language through a bottleneck is evolutionary more stable than transmitting holistic languages

because the dynamics of cultural evolution in vertical transmission – as in the iterated learning model – is quite different from the dynamics that can be observed in systems pertaining to oblique and horizontal transmission, which are more reminiscent of human cultural evolution (Cavalli-Sforza and Feldman 1981). These systems allow for cultural traits, such as linguistic entities or memes, to evolve based on neo-Darwinian evolution in which variation, competition and self-organisation of traits play a crucial role (Boyd and Richerson 2005; Croft 2002; Mufwene 2001). One advantage of a transmission system where the offspring can (try to) transmit knowledge to peers or to older generations while they are still learning, is that they will encounter new situations in which they may need to communicate about previously unseen items (cf. Fig. 8.1). In the iterated learning model such situations only occur after learning has stopped. The system of horizontal and oblique transmission thus provides learners with a natural implicit transmission bottleneck that triggers the emergence of compositionality (Vogt 2005c).

A downside of predefining the agents' semantics – as is the case in most iterated learning models – is that this removes (1) the role that ontogenetic development of concepts can play in bootstrapping the emergence of compositionality (Vogt 2006b), and (2) the individual variation in conceptualisation which is a crucial component of neo-Darwinian evolution. Moreover, enabling agents to develop categories/concepts from interacting (i.e. perceiving and acting in) with the world, it becomes important to consider by what means the world is perceived and acted in. For example, a researcher should consider what sensors a robot may have. Are these cameras, touch sensors, a compass or a combination of these? And what type of information is filtered from these sensors? All these factors essentially define the agents' embodiment, which in turn defines what qualities the agent can perceive in the world, thus constricting the possible conceptual spaces that can be formed. Although agents with different physiological capacities can learn

to communicate effectively – think of blind people, but also robots can do this (de Greeff and Belpaeme 2011) – the question is to what extent they converge on internal conceptual representations. This question is even relevant for agents having the same bodies, but different experiences in the world.

It should be clear that agents form concepts that reflect the world they engage in – it is impossible for agents who only encounter a flat world to acquire concepts such as uphill or downhill. People who only live in a remote area of the Amazon and who have never visited or seen skyscrapers, will not be able to fully grasp the concept of a skyscraper. This does not only apply to basic concepts, but also to compositions of concepts and the structures thereof. For instance, consider the concept of a cup. A cup can hold many substances (coffee, tea, water, sugar, ...), have various colours (white, blue, orange, ...), shapes, textures, sizes, etc. When there is only one cup present in a particular context, such features are not so important, but when there are multiple cups around these features may become important. The way humans conceptualise the cup in these different situations is hard to tell, but looking at the ways humans refer to a particular cup in different situations suggest we structure our conceptual representation (Brennan and Clark 1996; Koolen et al. 2011). The way concepts are structured depends strongly on the objects' properties and the way we perceive them, which in turn tends to be reflected in the language. I would argue that this goes so far that much of the structure of our engagement in the world (and more particularly in our ecological niche) is reflected in the grammars of our language. Humans tend to manipulate some target in one way or another. This is how we universally behave in the world, and that is what is reflected in most languages spoken across the globe: Most (but not all) languages have linguistic structures in which sentences contain a subject, a verb and an object (Baker 2003; Evans and Levinson 2009). Hence, the way we interact with our environment (i.e. our situatedness) and consequently the structure of our environment, as well as our embodiment, influence the way we conceptualise the world. Culture and language are part of our environment and are thus not only manifestations of our conceptualisations, but also shape them.

In the remainder of this chapter, I discuss a model that tries to incorporate the aforementioned principles in a simulation in which a population of agents evolve a simple compositional language from scratch in two steps: first a holistic language is formed, second a transition towards a compositional language occurs (Vogt 2005a,c, 2007). This model combines some components of Kirby's iterated learning model (Kirby 2001) – language learning and transmission over generations – with Luc Steels' language game model (Steels 1997, 2003, 2012). This way, grammatical structures and – as part of this – conceptual spaces co-evolve through self-organisation driven by social interactions between agents and the cognitive learning mechanisms of these agents. As the agents are situated in a virtual environment where they are forced to communicate about the objects in the environment, the structure of the environment, as well as the agents' perceptual apparatus, constrain the conceptual structures of the emerging languages. The general principles of this system – especially with regards to the complex adaptive dynamics – are the same as in most of Steels' studies. However, where the formation of grammar in Steels'

models relies on a complicated formalisation of cognitive grammars (Steels and Beule 2006; Steels 2012), the model presented here relies on a straightforward realisation of alignment-based learning (van Zaanen 2000) in combination with data-oriented parsing (Bod et al. 2003).

8.3 Language Games

The model simulates the Talking Heads experiment (Steels et al. 2002) in which a population of agents play a large number of guessing games – a variant of the language game – to develop a language that allows the population to communicate about their world. This world contains 120 coloured geometrical shapes (12 colours $\times$ 10 shapes) and the agents can only perceive the RGB values of the colour and one feature representing the shape. A guessing game is played by two agents: a speaker and a hearer. The aim of the game is for the hearer to guess what the speaker verbally refers to, and – where possible – each individual agent adapts its conceptual and linguistic representations such that the communication becomes more effective. The game consists roughly of the following steps: perception, conceptualisation, production, interpretation and adaptation.

These steps are explained in some detail in the remainder of this section, with a special focus on the emergence of conceptual spaces. It is beyond the scope of this chapter to present all details of the model, and the interested reader is referred to Vogt (2005a,c).

8.3.1 Perception and Conceptualisation

In each guessing game, a number of objects are randomly drawn from the world with a uniform distribution and are 'shown' to the agents as the context of the game. Suppose an agent sees the three objects on the top left of Fig. 8.2: red square, yellow hexagon and purple circle. Using its perceptual apparatus, each object is transformed into a 4-dimensional vector representing the r, g and b values of the RGB colour space and a feature value representing the object's shape s. The red square is thus represented by vector $(1, 0, 0, 1)$, the yellow hexagon by $(1, 1, 0, 0.5)$ and the purple circle by $(1, 0, 1, 0.57)$. These feature vectors represent the raw percepts of the objects.

Each feature of each percept is then categorised with a category from the relevant r, g, b and s quality dimensions. The categories divide each dimension in one or more segments and are represented by a prototypical value, as indicated by a dot in Fig. 8.2. The square would be thus categorised by the set $\{A, E, F, G\}$, the hexagon by $\{A, D, F, G\}$ and the circle by $\{A, E, F, G\}$. Such sets represent the objects' concepts as cubes in the 4-dimensional conceptual space. This is, probably, not a realistic representation of conceptual spaces, but it is a consequence of treating each

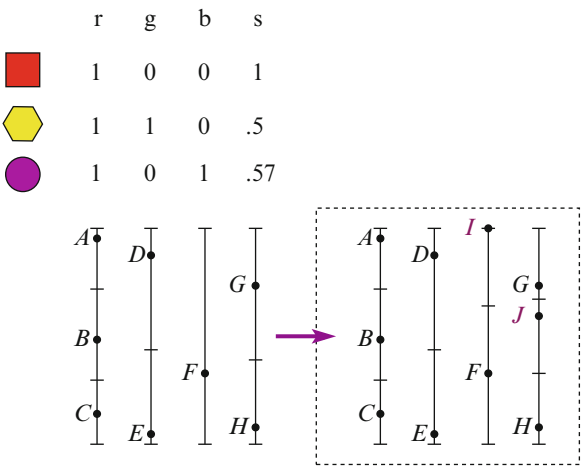

Fig. 8.2 This figure illustrates the conceptualisation and adaptation within the discrimination game (see the text for details)

quality dimension independently to facilitate their selection to be part of different conceptual spaces. More realistic implementations of conceptual spaces that would be applicable have been put forward in, e.g., Steels and Belpaeme (2005), Wellens et al. (2008) and Vogt (2004).

In order to communicate effectively, the agents individually process discrimination games (Steels 1996). The object of a discrimination game is to obtain a concept that represents an object such that it distinguishes this object from the other objects in the context. If the agent is to conceptualise the hexagon in contrast to the two other objects of Fig. 8.2, the agent is successful as the concept $\{A, D, F, G\}$ is distinctive. If, however, the agent is to discriminate the square or the circle from the other two objects in this context, the agent fails as both objects have the same concept. If this occurs, the agent adapts its categories by adding the feature values of distinguishable dimensions as a prototypical exemplar to the appropriate quality dimensions. For instance, if the agent was trying to distinguish the circle from the other two objects, it would add the categories I and J to the original representation, yielding the set of quality dimensions with categories as depicted in the dashed box of Fig. 8.2.

Conceptual spaces in this model can be formed by taking one to four of these quality dimensions together, so there can be four 1-dimensional spaces, six 2-dimensional spaces, four 3-dimensional spaces and one 4-dimensional space. The concepts within each space can be used to represent the basic meanings in the agents' language. This way, conceptual spaces are constructed that could be interpreted as linguistic categories. Initially, the agents will only conceptualise percepts in the 4-dimensional space and associate such concepts with word-forms in a holistic manner. The purpose of this study is to demonstrate how agents can develop conceptual spaces of lower dimensions and use these coherently in language. To understand how this may be achieved it is important to understand how the agents represent, use and learn their language.

8.3.2 Production and Interpretation

Once the agents have categorised the objects in the context, the speaker selects one object at random with a uniform distribution as the topic of the communication. This agent then searches its grammar for ways to produce an expression that conveys the topic's concept. The grammar (Fig. 8.3) is an individual's competence and consists of simple rewrite rules that associate forms with concepts either holistically (e.g., rule 1) or compositionally (e.g., rule 2 combined with rules 3 and 4). The grammar may be redundant in that there may be rules that compete to produce or interpret an expression (cf. Batali 2002; De Beule and Bergen 2006; Steels 2012). The speaker searches for those (compositions of) rules that match the topic's concept and if more than one are found, he selects the rule that has the highest rule score. If the speaker fails to produce an expression this way, a new form is invented as an arbitrary string and is associated with the topic's concept or – if a part of the concept matches some non-terminal rule – with the complement of this concept. For instance, if the speaker would want to produce an utterance expressing a red square (1, 0, 0, 1) and it knows a word for the colour red (1, 0, 0, ?) but not for square, then it invents a new word (e.g., 'wateva') to express square (?, ?, ?, 1) and adds this to its grammar.

In turn, the hearer tries to interpret the expression by searching its own grammar for (compositions of) rules that match both the expression and a concept relating to an object in the current context. If there is more than one such rule, the hearer selects the one with the highest score, thus guessing the object intended by the speaker. The hearer then 'points' to this object, and if this is the object intended by the speaker, the speaker acknowledges success; otherwise, the speaker points to the topic allowing the hearer to acquire the correct concept referring to the expression.

1	S → greensquare/(0,1,0,1)	0.2
2	S → A/rgb B/s	0.8
3	A → red/(1,0,0,?)	0.6
4	B → triangle/(?,?,?,0)	0.7

Fig. 8.3 This example grammar contains rules that rewrite a non-terminal into an expression-meaning pair (1, 3 and 4) or into a compositional rule that combines different non-terminals (2). Rule (2) is thus a rule that combines linguistic categories/conceptual spaces A/rgb and B/s (i.e., A relates to the RGB colour space and B to the shape space). For practical reasons concepts are presented as 4-dimensional vectors, where the first 3 dimensions relate to the RGB colour space (rgb) and the 4th relate to the shape feature (s); the question marks are wild-cards and indicate which quality dimension(s) is (are) not part of this conceptual space. Each rule has a rule score that indicates its effectiveness in past guessing games. Only sentences of one or two constituents are allowed in this grammar

8.3.3 Adaptation

If the guessing game was successful, both the speaker and hearer increase the scores of the rules they used and lower the scores of those rules that compete with the used rules. If the game has failed, the scores of used rules are lowered and the hearer acquires the proper association between the heard expression and the topic's concept. To this end, the hearer tries the following three steps until one step has succeeded:

1. If a part of the expression can be interpreted with a part of the topic's concept, the rest of the expression is associated with the complement of the concept. For instance, if the hearer of the grammar shown in Fig. 8.6 hears the expression "redcircle" referring to the concept (1,0,0,.5), the part "red"-(1,0,0,?) can be interpreted, so the hearer adds rule B→circle/(?,?,?,.5) to its grammar.
2. If the above failed, the hearer searches its memory, where it stores all heard or produced expression-concept pairs, to see if there are instances that are partly similar to the expression-concept pair just heard. If some similarity can be found, the hearer will break-up the expression-concept pairs containing these similarities following certain heuristics, thus forming new compositional rules. Suppose, for instance, the hearer had previously heard the expression-concept pair "greensquare"-(0,1,0,1) and now hears "yellowsquare"-(1,1,0,1). The hearer can then break up these pairs based on the similarity "square"-(?,1,0,1), thus forming rules S→C/r D/gbs, C→green/(0,?,?,?), C→yellow/(1,?,?,?) and D→square-(?,1,0,1). Note that this is not the ideal break up, since it breaks apart the red component of the RGB colour space from the blue and green components and the shape feature (3). The next section shows that over time such mistakes diminish as a result of competition and selection.
3. If the above adaptations both fail, the heard expression-concept pair is incorporated holistically, leading to a new rule such as S→yellowcircle/(1,1,0,.5).

At the end of these steps, the hearer performs a few post-processes to remove any multiple occurrences of rules and to update the grammar such that other parts of the internal language relates more consistently to the new knowledge. Full details of the model are found in Vogt (2005a,c).

The three learning steps are the core cognitive mechanisms responsible for the co-evolution of linguistic structures and conceptual spaces. Basically, if there is no compositional structure yet in the rules of an agent, but there are regular patterns (i.e. similarities) in both forms and concepts, they are both split up. Yet, this does not necessarily mean these new rules will survive in the language. The way an agent breaks apart holistic expression-concept pairs depends on what the agent has acquired before, so it may make errors. However, later on in life the agent can recover from these errors when it hears new and different usages of parts of an expression. When that occurs, the agent adds new variants to its 'pool' of transmissible information units, which then compete for being used. Elements from these pools are selected based on their effectiveness in communication. If an

element is used ineffectively, it is dampened and when it is used effectively it is reinforced, while competing ones are laterally inhibited. This competition yields a self-organising effect on the languages of the individual agents, but also brings about effectiveness at a global level, such that a globally shared language can evolve.

In the model, agents have four quality dimensions at their disposal and initially recruit them to form the conceptual space holistically. During development when the holistic expression-concept pairs are broken apart, the agents form new linguistic categories, each semantically relating to a conceptual space of lower dimensionality. The cognitive mechanism for breaking apart expression-concept pairs does not only require an alignment in expressions, but also in conceptual representations. This way a co-evolution of language and concept emerges that on the linguistic side is driven by cultural transmissions and on the conceptual side is facilitated and constrained by the environment (i.e. the objects in the world) and embodiment (i.e. the categorisation into quality dimensions). These processes are all mediated (i.e., facilitated and constrained) by the cognitive capacities of the agents.

8.4 Simulating the Evolution of Conceptual Spaces

In order to illustrate the framework described in the first part of this chapter and to illustrate the conditions in which a compositional structure of conceptual spaces can emerge, two simulations were carried out. The first simulation, previously reported in Vogt (2006a), illustrates how the model evolves to a sub-optimal solution when there is no generational turnover, so where there is only horizontal transmission. The second simulation demonstrates that more optimal solutions emerge when there is a population flow such that the population contains multiple generations.

Before presenting the results, two measures need to be defined:

Communicative success measures the number of successful guessing games over a time window of 50 games.

Similarity measures the number of games in which both agents used the same syntactic structure over a time window of 50 games. A syntactic structure is considered similar if the words and the linguistic categories used are the same and in the same order. (A linguistic category is characterised by the dimensions that make up the conceptual space of a non-terminal node.)

Both measures are normalised to a value between 0 and 1. Communicative success informs us how successful the population becomes in communicating the referents. This measure, however, does not inform us how similar the internal languages are – the agents may well use different representations and nevertheless be successful in communication. Similarity informs us about the extent in which agents use the same grammatical constructions, thus to what extent they use the same conceptual spaces.

To show the evolution of conceptual spaces in more detail, I also present the relative frequencies of rule types used during successive periods of 10,000 guessing

games. As the agents can break up the 4-dimensional conceptual space into two conceptual spaces of lower dimensions without having prior knowledge which dimensions should be separated, 15 different rule types (including the holistic type) can emerge. Only 5 rule types are inspected in this chapter (all other had very low frequencies):

I:	S→rgbs	holistic rule
II:	S→A/r B/gbs	red v. green, blue & shape
III:	S→B/gbs A/r	green, blue & shape v. red
IV:	S→C/rgb D/s	colour v. shape
V:	S→D/s C/rgb	shape v. colour

Rule type I concerns holistic rules in which word forms are associated with the 4-dimensional conceptual space. Rule types II and III are rules that combines the 1-dimensional conceptual space of the quality dimension that represents the red component of the RGB space with the 3-dimensional conceptual space containing the quality dimensions representing the green and blue RGB components, and the shape dimension. The difference between the two rule types is word-order. Rule types IV and V combines the 3-dimensional conceptual space that represents colour in the RGB space with the 1-dimensional shape space.

One of the reasons for inspecting rule types II and III is that in this world, the probability of finding a regularity in the red component of the RGB space is substantially higher than finding any other regularity, such as those required to establish rules IV and V (Table 8.1). The probability of finding a regular pattern in the RGB space versus the shape space (cf. rule types IV and V) between two randomly selected objects is the chance that the two objects have the same colour (1/12) times the chance that the two objects have different shapes (9/10), which thus becomes $1/12 \cdot 9/10 = 0.075$. The probability of finding a regular pattern in the red component is much higher, because the 12 colours used in the simulation are highly regular in this dimension: 4 colours have value 0, 5 have value 1 and the others have unique values. Without showing the exact calculation, the average probability of finding a regularity in the red component of two randomly selected objects, while the values of the other dimensions differ, is 0.297. The probability of finding regularities in combinations of other dimensions (e.g. g-rbs, b-rgs, rg-bs, etc.) is somewhere in between (cf. Table 8.1). Although the rules for these combinations would occur more frequently by chance than rules of types IV and V, these are seldomly used by the agents, so their occurrences are not presented.

Despite the probability of finding a regularity in the red component is highest, rule types II and III which exploit this component are not efficient in terms of grammar size. This is because the complements of the red component in the RGB space are not very regular. In fact, the 12 colours have 9 different complements composed by the green and blue RGB components (three of which occur twice, both with red component values of 0 and 1). When combined with the 10 different

Table 8.1 The probability P of finding in two different games a co-occurring structure in conceptual space X and not in Y in which case the 4-dimensional space may be segmented into these two spaces. These probabilities are based on the distribution of feature values that represent the different objects in the world (This table is reproduced from Vogt (2005b))

X	Y	P
r	gbs	0.297
g	rbs	0.200
b	rgs	0.256
rg	bs	0.117
rb	gs	0.144
gb	rs	0.117
rgb	s	0.075

shapes, the grammar to describe all 120 coloured shapes, would contain at least 96 rules: 5 to cover the red component, 90 to cover the gbs-space and one to describe word-order. In contrast, rules of type IV and V (i.e. those that combine colour with shape) only require a grammar of 23 rules: 12 to cover the rgb-space, 10 to cover the s-space and one to describe word-order. Thus, the two rule-types combining colour with shape are most optimal in terms of compressibility.

8.4.1 Horizontal Transmission

The first simulation is the same as the one reported earlier in Vogt (2006a), but now discussed in the light of the framework set out earlier. This simulation involves a population of 50 agents from the same generation and is run for 1 million guessing games. In each game two agents are selected at random, one agent is arbitrarily assigned the role of speaker, and the other the role of hearer. The context size in each game was set to eight objects, randomly drawn from the world of 120 objects without replacement. Previous research has shown that there is little variation in the results when the simulations are replicated 10 times with different random seeds (Vogt 2005a,c). For the purpose of this chapter it is instructive to look at the results from one simulation run.

Figure 8.4 shows the results of a typical simulation. The top graph shows that communicative success rapidly increases to a value near 0.5, after which it slowly increases to a value slightly above 0.8 and after around 500,000 guessing games, the system stabilised and more than 80 % of the games are successful. Similarity (middle graph), however, increases to a value around 0.5, after which it stops increasing. So, in nearly half of the games, the agents use different internal grammar rules, even if they use the same utterances to refer to an object successfully. For example, some agents may use a holistic rule (type I), while others use may rule type II, III, IV or V.

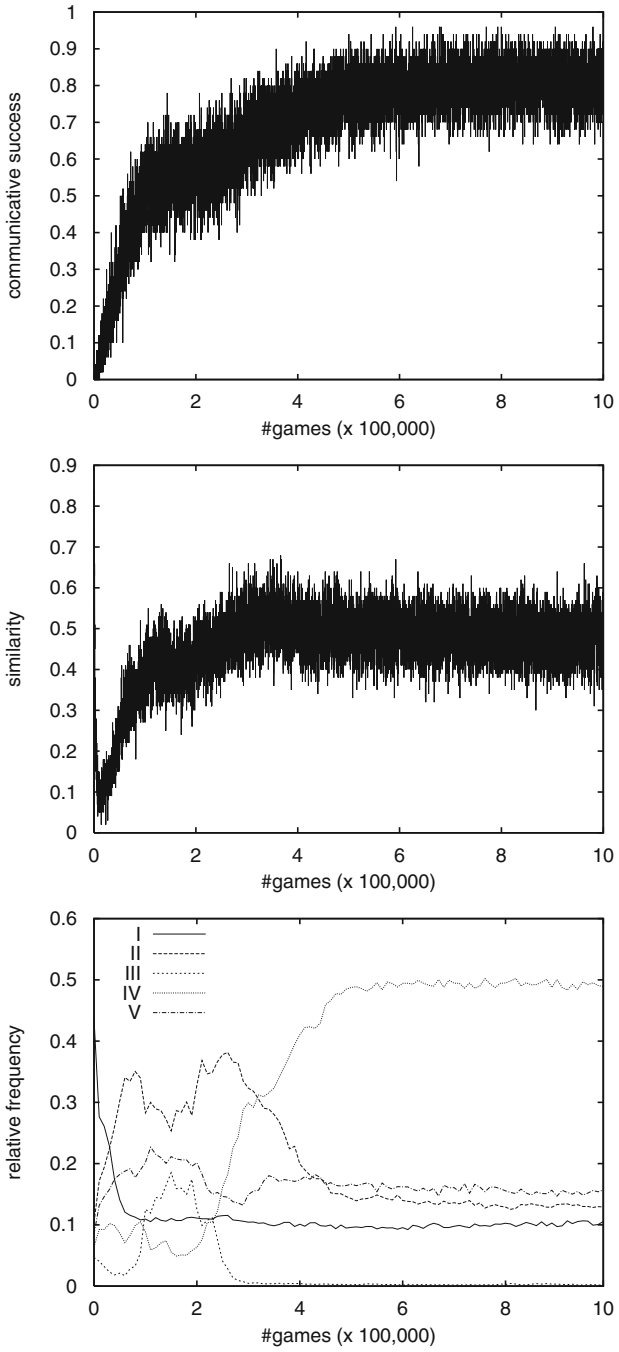

Fig. 8.4 The results of the first simulation. The graphs show communicative success (*top*), similarity (*middle*) and the competition diagram showing the evolution of rule types (*bottom*) (These figures are reprinted with permission from Vogt (2006a))

The competition diagram (Fig. 8.4, bottom) shows the relative frequencies of the five rule types during this simulation. In the first 200,000 games, all rules types compete to be used. At the very early stages, the holistic rule (type I) occurs most frequently, but soon drops to a value near 0.2 after which it stabilises. So, in about 10 % of the interactions, the agents use the 4-dimensional conceptual space to communicate objects. The other 90 % are divided among all other rule types (including those not shown). After a bit more than 200,000 games, the frequency of rule type III drops to a value near 0, while rule types II and IV appear to compete for some more time until the system more or less stabilised after 500,000 games. From this time onward, the most frequently used rule type is number IV, followed by rule types V, II and I respectively.

So, although communicative success is high, similarity in the representation of the individual grammars (and consequently conceptual spaces) as used by the different agents has evolved into a sub-optimal system. About 70 % of the rules used by the agents depend on conceptual spaces rgb and s (rule types IV and V), about 15 % by conceptual spaces r and gbs (rule type II), and 10 % by the 4-dimensional conceptual space (rule type I). This is sub-optimal, because the most efficient way of representing the grammars is by using rule types IV and/or V, since these require the least number of rules to capture the entire world.

8.4.2 Isotropic Transmission

In the second simulation, the same model was used with the same parameter settings, but, instead of having one generation to simulate horizontal transmission, this simulation implements a more naturalistic population flow (cf. de Boer and Vogt 1999; Steels and Kaplan 1998). By allowing all agents speak to all other agents, this system implements *isotropic transmission* (Vogt 2005c) that combines oblique, horizontal and upward forms of transmission. To implement a population flow, each agent was given an age measured in terms of the number of guessing games they played individually with a maximum set to 12,500 games. As before, the simulation starts with 50 agents, each initialised with an arbitrary age between 0 and 12,500 games. Each time an agent has played a game, this agent could die with a Gaussian probability distribution with the mean set to the maximum age and a standard deviation of 250. When one agent thus dies, a new agent is added to the population to keep the population size fixed at 50. New agents start with neither concepts nor grammar. Figure 8.5 shows the total number of agents that have entered the simulation over time.

Figure 8.6 shows the results of this simulation. The first thing that should be noted is that, in addition to the spikes, there are more fluctuations in the trends of the different graphs. These fluctuations coincide with increased influx of agents as shown in Fig. 8.5. Apart from these fluctuations, it is apparent that communicative success rises to a similar level as in the previous simulation, but similarity rises to a substantially higher level and settles to fluctuate around 0.75. So, agents

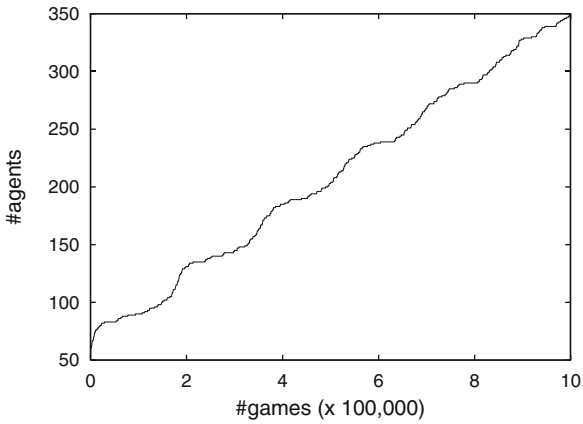

Fig. 8.5 The total number of agents that have entered the system over time

increasingly agree on using the same rule types. In particular, from about 300,000 games onward rule type V is most frequent (about 70 %), followed by rule type IV (about 20 %). The holistic rule type I continues to decrease from around 10 % at 300,000 games to less than 5 % at the end. The other rule types are only sparsely used.

These results demonstrate that when there is a generational turn-over, the language and conceptual spaces continue to evolve towards an optimal system where the grammar represents rules that combine colour with shape in slightly more than 90 % of all cases, rather than stabilising in a sub-optimal system as in the previous simulation. So, the new agents rapidly learn the established language by acquiring and using the optimal rule types more effectively than the other rules.

8.5 Discussion

The simulations presented in the preceding section demonstrated how different conceptual spaces can emerge through cultural evolution. As argued in Sect. 8.2, the following factors are involved in this evolution: the environment, embodiment, cognition, self-organisation and cultural transmission. The remainder of this chapter will discuss how these factors contribute to the observed evolution in the simulations, starting with the first three factors, because they are highly interrelated.

The environment of the agents consists of objects that combined a given set of primary colours with a given set of basic shapes. As such, the most obvious way of expressing (and hence conceptualising) these objects is by colour and shape. The agents were embodied with feature detectors that represent the three dimensions of the RGB colour space and one detector that gives a value for each object, however, the agents had no way of telling which of these feature detectors belong to colour or shape and were treated independently. The quality dimensions these agents were endowed with constrained the way they could categorise the perceived

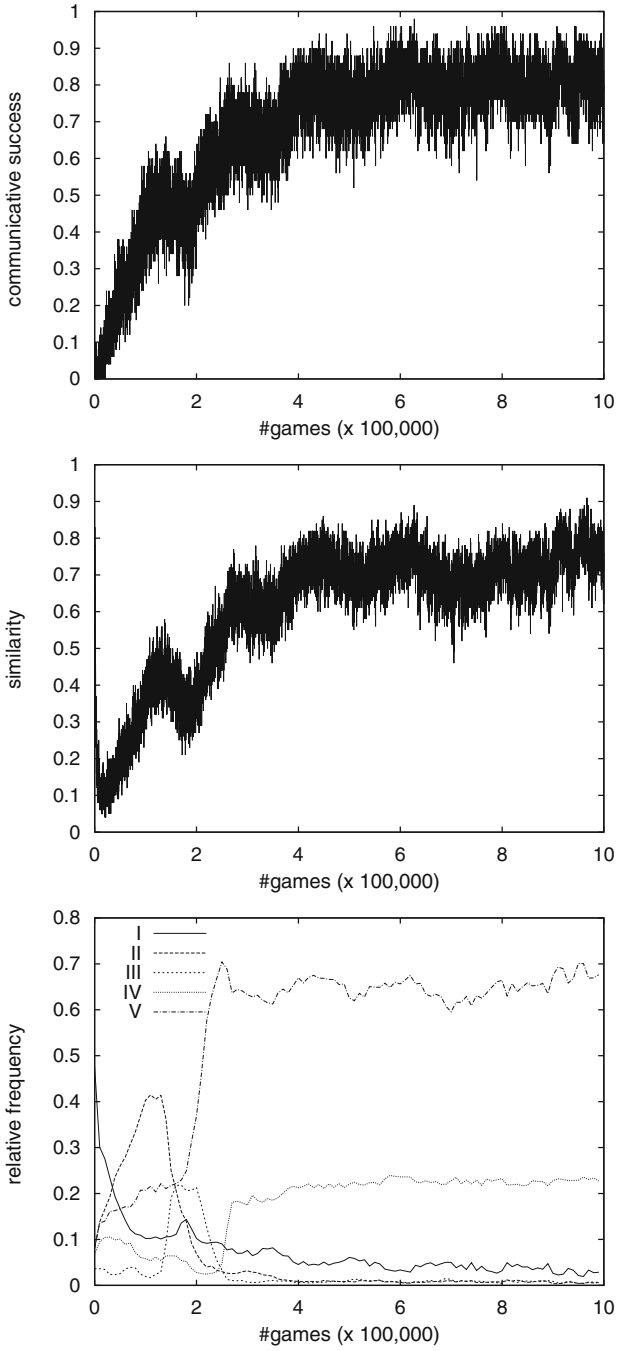

Fig. 8.6 The results of the second simulation. The graphs show communicative success (*top left*), similarity (*bottom left*) and the competition diagram showing the evolution of rule types

objects and how these could be combined to form different conceptual spaces. The cognitive mechanisms were designed such that the agents could only acquire and use grammatical rules that either treated the semantics to be represented holistically or as a combination of two conceptual spaces with the restriction that all quality dimensions are used exactly once. As a result, the agents could construct a total of 15 different conceptual spaces to be used in 8 different combinations irrespective of word-order.

The way the colours and shapes were constructed to form the environment and the way agents could perceive these determined the distribution of focal values in each quality dimension. Since the way agents induce compositional rules from the observed input is based on discovering regular aligned patterns in two or more utterance-concept pairs (as outlined in Sect. 8.3.3), the probabilities of finding a regular pattern would drive the formation of grammatical rules as shown in Table 8.1. To some extent, this is observed in the simulations where rule types II and III occur frequently (at least in the beginning), but all other compositional rule types, except types IV and V, were hardly used. The explanation for this relates to an interaction between the environment, the cognitive learning mechanism and self-organisation.

The environment was constructed such that despite the probabilities of finding regular patterns in all combinations, except colour and shape, were higher, the combination of colour and shape would yield the most compact grammar to express the world. The utilisation of this property would not have happened without the feedback loop – the reinforcement of rule scores and the resulting self-organisation. When agents receive positive feedback, they increase the scores of rules that were used. In cases where agents have different ways of expressing an object by using different combinations of rules, they will select those rules that have the highest combined scores. Since the rules combining colour and shape could apply for all objects, these rule types are more likely to be reinforced and thus more likely to be re-applied. When these rules are more frequently re-applied by the speaker, this increases the chance that the hearer would discover a regular pattern in colour and shape. This positive feedback loop is a driving factor of self-organisation, similar to the way ant paths are formed (Prigogine and Strengers 1984), and is considered one of the strongest factors for convergence in the language game paradigm (Steels 1997). Although in the first simulation language evolved into system that incorporated rules combining colour and shape most frequently, a substantial amount of rules of types I and II remained. The constructions formed with these rules were so entrenched in the language that they were viable, also because the language had evolved into a stable system with no variation.

Variation, which is one of the crucial ingredients of (neo) Darwinian evolution (Boyd and Richerson 2005; Darwin 1968; Dawkins 1976) and which is thought to be a driving factor of language change (Croft 2000; Mufwene 2001), occurs in the system through the speakers' invention of new words and through the acquisition of new constructs by hearers. In the first simulation, all variant constructions are created and spread among the population in the first, say 100,000 games or so, and after that the competition between the variants take over, which after approximately

500,000 games yield a stable system. The initial variation, subsequent competition and evolution to a stable system is characteristic of the language game model as is most clearly demonstrated in the naming game studied by Baronchelli et al. (2006). When, as in the second simulation, newborn agents continue to enter the population and learn the language from scratch, the system no longer gets stuck in such a sub-optimal stable system.

The reason for a continued evolution is that these young agents create new variations in the pool of utterances. Often these new variations are errors or over-extensions (Vogt 2006b) that tend to be unlearnt during development, but sometimes these are new variations introduced by applying a compositional rule to previously unseen objects (a result of the implicit bottleneck, see Section 2 and Vogt 2005c). Since the rules combining colour and shape tend to occur most frequently in the language (see competition diagrams of Figs. 8.4 and 8.6), it is most likely that these new variants reflect that structure. As a result, even more utterances that comply to these rules enter the language, increasing the chance for other agents to discover and use those regularities. This cultural transmission over generations thus strengthens the positive feedback loop that drives the self-organisation. Both language and concepts thus co-evolve to be learnt easier, as there are less rules to acquire (cf. Kirby and Hurford 2002).

It is important to note that due to the – necessary – abstractions made in this model, it is hard to generalise the results from this study to the way humans form conceptual spaces. The simulations are situated in a toy world, with homogeneous agents who can perceive the objects identically and without noise. In addition, the way concepts are constructed from independent quality dimensions is probably unrealistic. Moreover, the model assumes that during the course of human language evolution, protolanguages were essentially holistic and gradually evolved into compositional languages. This assumption is still very much under debate (Arbib and Bickerton 2010). In spite of these abstractions, the model also contains a number of more realistic assumptions, such as a gradual generational turnover in the population, mechanisms that facilitate self-organisation, and general mechanisms for detecting regularities in the input. As a result, the present study illustrates plausible theoretical principles that may explain how conceptual spaces are shaped. Future modelling work should investigate the scalability of this model using a more realistic world (perhaps even the real world) and agents with more human-like like embodiment and cognition.

8.6 Conclusions

This chapter has investigated how conceptual spaces can emerge from quality dimensions based on the cultural evolution of compositional languages. The same principles have been demonstrated before in a series of studies where the population flow was implemented based on the iterated learning model in which the population always contained two generations (adults and children) and after a predetermined number of games, all adults die, children become adults and new children enter the

population (see Vogt 2007, for an overview). The differences between the present study and those previous studies concern the more gradual population flow and the focus on conceptual spaces.

The simulations have demonstrated that the evolution of conceptual spaces is driven by five crucial factors: environment, embodiment, cognition, self-organisation and cultural transmission. The emerging conceptual spaces reflect the structure of the environment. Its development within the agents is facilitated by the embodiment through its perceptual apparatus and the cognitive mechanisms. However, embodiment and cognition (and arguably the environment as well) are at the same time limiting factors. Would the agents have been able to perceive other qualities or to manipulate objects, then more complex languages could have evolved, provided the cognitive learning mechanisms would allow them to break apart the holistic utterances in more than two constituents.

The self-organisation results from the variation and competition in conceptual and linguistic structures, as well as the positive feedback loop driven by the learning mechanism. Cultural transmission across generations allows for additional variations to prevent the system entering a sub-optimal stable system and keep the evolution going. Gradually, the emerging language becomes easier to learn, which can catalyse cumulative cultural evolution (Boyd and Richerson 2005; Vogt 2006a). Due to the limitations that the model imposed on environment, embodiment and cognition, the linguistic structures and consequently the conceptual spaces evolved into a relatively stable state. However, if there was room for further development, more complex structures could have emerged.

Crucial to the design of this is the assumption that language and concepts co-evolve. This is in line with the renewed appreciation of Whorf's linkage between language and thought (Bowerman and Levinson 2001), and which may account for the cross-cultural differences in the ways languages express and conceptualise various aspects of the world, such as spatial relations (Haun et al. 2011; Majid et al. 2004). Although the present study did not focus on cultural differences in conceptualisation, the framework has the potential to explain these. To achieve this, future studies should incorporate more realistic scenarios based on data from different cultures, as for instance collected by Vogt and Mastin (2013).

Acknowledgements The writing of this chapter was funded through a Vidi grant provided by the Netherlands Organisation for Scientific Research (NWO, grant no. 276-70-018). I wish to thank Frank Zenker, Peter Gärdenfors and all participants of the Conceptual Spaces at Work symposium for their valuable contributions in discussing this research. Also, many thanks to Emiel Krahmer and an anonymous reviewer for their valuable comments on earlier versions of this manuscript.

References

Arbib, M. A., & Bickerton, D. (Eds.). (2010). *The emergence of protolanguage: Holophrasis vs compositionality*. Amsterdam: John Benjamins.

Baker, M. C. (2003). Linguistic differences and language design. *Trends in Cognitive Sciences, 7*(8), 349–353.

Baronchelli, A., Felici, M., Caglioti, E., Loreto, V., & Steels, L. (2006). Sharp transition towards shared lexicon in multi-agent systems. *Journal of Statistical Mechanics 2006*, P06014.

Batali, J. (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (chap. 5). Cambridge: Cambridge University Press.

Bickerton, D. (1984). The language bioprogram hypothesis. *Behavioral and Brain Sciences, 7*, 173–212.

Bod, R., Sima'an, K., & Scha, R. (Eds.). (2003). *Data oriented parsing*. Stanford: Center for Study of Language and Information (CSLI) Publications.

Bowerman, M., & Levinson, S. C. (Eds.). (2001). *Language acquisition and conceptual development*. Cambridge: Cambrige University Press.

Boyd, R., & Richerson, P. (2005). *The origin and evolution of cultures*. Oxford: Oxford University Press.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1482–1493.

Brighton, H., & Kirby, S. (2001). The survival of the smallest: Stability conditions for the cultural evolution of compositional language. In J. Kelemen & P. Sosík (Eds.), *Proceeding of the 6th European conference on artificial life, ECAL 2001*, Prague.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton: Princeton University Press.

Croft, W. (2000). *Explaining language change: An evolutionary approach*. New York: Longman.

Croft, W. (2002). The Darwinization of linguistics. *Selection, 3*, 75–91.

Darwin, C. (1968). *The origin of species*. London: Penguin Books.

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

De Beule, J., & Bergen, B. K. (2006). On the emergence of compositionality. In A. Cangelosi, A. Smith, & K. Smith (Eds.), *The evolution of language: Proceedings of the 6th international conference on the evolution of language*, Rome.

de Boer, B., & Vogt, P. (1999). Emergence of speech sounds in changing populations. In D. Floreano, J.-D. Nicoud, & F. Mondada (Eds.), *Advances in artificial life: Proceedings of 5th European conference ECAL'99*, Lausanne.

de Greeff, J., & Belpaeme, T. (2011). The development of shared meaning within different embodiments. In *2011 IEEE international conference on development and learning (ICDL)*, Frankfurt am Main (Vol. 2, pp. 1–6).

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*(05), 429–448.

Gärdenfors, P. (2000). *Conceptual spaces*. Cambridge, MA: Bradford Books/MIT.

Garrod, S., Fay, N., Rogers, S., Walker, B., & Swoboda, N. (2010). Can iterated learning explain the emergence of graphical symbols? *Interaction Studies, 11*(1), 33–50.

Haun, D., Rapold, C. J., Janzen, G., & Levinson, S. C. (2011). Plasticity of human spatial cognition: Spatial language and cognition covary across cultures. *Cognition, 119*(1), 70–80.

Jackendoff, R. (2002). *Foundations of language*. New York: Oxford University Press.

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin and Review, 14*(2), 288–294.

Kirby, S. (2001). Spontaneous evolution of linguistic structure: An iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation, 5*(2), 102–110.

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121–148). London: Springer.

Kirby, S., Smith, K., & Brighton, H. (2004). From UG to universals: Linguistic adaptation through iterated learning. *Studies in Language, 28*(3), 587–607.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105*(31), 10681–10686.

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics, 43*(13), 3231–3250.

Majid, A., Bowerman, M., Kita, S., Haun, D., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences, 8*(3), 108–114.

Mufwene, S. S. (2001). *The ecology of language evolution*. Cambridge: Cambridge University Press.

Pinker, S., & Bloom, P. (1990). Natural language and natural selecion. *Behavioral and Brain Sciences, 13*, 707–789.

Prigogine, I., & Strengers, I. (1984). *Order out of chaos*. New York: Bantam Books.

Steels, L. (1996). Perceptually grounded meaning creation. In M. Tokoro (Ed.), *Proceedings of the international conference on multi-agent systems*, Kyoto. Menlo Park: AAAI Press.

Steels, L. (1997). Synthesising the origins of language and meaning using coevolution, self-organisation and level formation. In J. Hurford, C. Knight, & M. Studdert-Kennedy (Eds.), *Approaches to the evolution of language*. Cambridge: Cambridge University Press.

Steels, L. (2003). Evolving grounded communication for robots. *Trends in Cognitive Sciences, 7*(7), 308–312.

Steels, L. (Ed.). (2012). *Experiments in cultural language evolution* (Vol. 3). Amsterdam/ Philadelphia: John Benjamins.

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences, 28*, 469–529.

Steels, L., & Beule, J. D. (2006). Unify and merge in fluid construction grammar. In P. Vogt, Y. Sugita, E. Tuci, & C. Nehaniv (Eds.), *Symbol grounding and beyond: Proceedings of the third international workshop on the emergence and evolution of linguistic communication*, Rome (pp. 197–223). Springer.

Steels, L., & Kaplan, F. (1998). Stochasticity as a source of innovation in language games. In *Proceedings of alive VI*. Los Angeles, CA.

Steels, L., Kaplan, F., McIntyre, A., & Van Looveren, J. (2002). Crucial factors in the origins of word-meaning. In A. Wray (Ed.), *The transition to language* (pp. 252–271). Oxford: Oxford University Press.

van Zaanen, M. (2000). ABL: Alignment-based learning. In *Proceedings of the 18th international conference on computational linguistics (COLING)*, Saarbrücken.

Vogt, P. (2004). Minimum cost and the emergence of the Zipf-Mandelbrot law. In J. Pollack, M. Bedau, P. Husbands, T. Ikegami, & R. A. Watson (Eds.), *Artificial life IX proceedings of the ninth international conference on the simulation and synthesis of living systems*, Boston (pp. 214–219). MIT.

Vogt, P. (2005a). The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence, 167*(1–2), 206–242.

Vogt, P. (2005b). Meaning development versus predefined meanings in language evolution models. In L. Kaelbling & A. Saffiotti (Eds.), *Proceedings of IJCAI-05*, Edinburgh (pp. 1154–1159). IJCAI.

Vogt, P. (2005c). On the acquisition and evolution of compositional languages: Sparse input and the productive creativity of children. *Adaptive Behavior, 13*(4), 325–346.

Vogt, P. (2006a). Cumulative cultural evolution: Can we ever learn more? In S. Nolfi et al. (Eds.), *From animals to animats 9: Proceedings of the ninth international conference on simulation of adaptive behaviour*, Rome. Berlin: Springer.

Vogt, P. (2006b). Overextensions and the emergence of compositionality. In A. Cangelosi, A. Smith, & K. Smith (Eds.), *The evolution of language: Proceedings of the 6th international conference on the evolution of language*, Rome. World Scientific Press.

Vogt, P. (2007). Variation, competition and selection in the self-organisation of compositionality. In B. Wallace, A. Ross, J. B. Davies, & T. Anderson (Eds.), *The mind, the body and the world: Psychology after cognitivism?* (pp. 233–256). Exeter: Imprint Academic.

Vogt, P., & Mastin, J. D. (2013). Anchoring social symbol grounding in children's interactions. *Künstliche Intelligenz, 27*, 145–151.

Wellens, P., Loetzsch, M., & Steels, L. (2008). Flexible word meaning in embodied agents. *Connection Science, 20*(2), 173–191.

Whorf, B. L. (1956). *Language, thought, and reality*. Cambridge: MIT.

Wray, A. (1998). Protolanguage as a holistic systemfor social interaction. *Language and Communication, 18*, 47–67.

Chapter 9
Logical, Ontological and Cognitive Aspects of Object Types and Cross-World Identity with Applications to the Theory of Conceptual Spaces

Giancarlo Guizzardi

Abstract Types are fundamental for conceptual domain modeling and knowledge representation in computer science. Frequently, monadic types used in domain models have as their instances *objects* (*endurants*, *continuants*), i.e., entities persisting in time that experience qualitative changes while keeping their numerical identity. In this paper, I revisit a philosophically and cognitively well-founded theory of object types and propose a system of modal logics with restricted quantification designed to formally characterize the distinctions and constraints proposed by this theory. The formal system proposed also addresses the limitations of classical (unrestricted extensional) modal logics in differentiating between types that represent mere properties (or *attributions*) ascribed to individual objects from types that carry a *principle of identity* for those individuals (the so-called *sortal types*). Finally, I also show here how this proposal can complement the theory of *conceptual spaces* by offering an account for kind-supplied principles of *cross-world identity*. The account addresses an important criticism posed to conceptual spaces in the literature and is in line with a number of empirical results in the literature of cognitive psychology.

9.1 Introduction

Types are fundamental for conceptual domain modeling and knowledge representation in computer science. Frequently, monadic types used in domain models have as their instances *objects*, i.e., entities that persist in time while keeping their identity (as opposed to *events* such as a kiss, a business process or a birthday party). What I term here *object* refers to what is sometimes termed Endurant or Continuant in the literature. Examples of objects include physical and social persisting entities of

G. Guizzardi (✉)
Computer Science Department, Federal University of Espírito Santo, Goiabeiras, Brazil
e-mail: gguizzardi@inf.ufes.br

F. Zenker, P. Gärdenfors (eds.), *Applications of Conceptual Spaces*,
Synthese Library 359, DOI 10.1007/978-3-319-15021-5_9

everyday experience such as balls, rocks, planets, cars, students and Queen Beatrix but also fiat objects such as the Dutch part of the North Sea and a non-smoking area of a restaurant.

In this paper, I revisit the philosophically and cognitively well-founded theory of object types first proposed in Guizzardi et al. (2004). The ontological distinctions and postulates proposed by this theory are discussed in the next section. In Sect. 9.3, I present the main contribution of this paper, namely, a system of modal logics with restricted quantification designed to formally characterize the distinctions and constraints proposed by this theory. That section also discusses how the proposed formal system addresses the limitations of classical (unrestricted extensional) modal logics in some fundamental aspects regarding the notions of object persistence and cross-world identity. In Sect. 9.4, the paper elaborates on how this theory can be employed to analyze and address some problems faced by the theory of *conceptual spaces* (Gärdenfors 2000) with respect to the issues of identity and persistence. Section 9.5 briefly discusses related work. Finally, Sect. 9.6 concludes the paper with final considerations.

9.2 Ontological Distinctions Among Object Types

Van Leeuwen (1991) presents an important grammatical difference occurring in natural languages between common nouns (CNs) and arbitrary general terms (adjectives, verbs, mass nouns, etc.. . .). Common nouns have the singular feature that they can be combined with determiners and serve as argument for predication in sentences such as: *(i) (exactly) five mice were in the kitchen last night; (ii) the mouse that ate the cheese has been in turn eaten by the cat.*

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

The explanation for this feature, which is unique of CNs, draws on the function that determinates (demonstratives and quantifiers) play in noun phrases, which is to determine a certain range on individuals. Both reference and quantification requires that the things that are referred or that form the domain of quantification are determinate individuals, i.e., their conditions for individuation and identity must be determinate.

According to van Leeuwen (1991), this syntactic distinction between the two linguistic categories reflects a semantical and ontological one, and so the distinction between the grammatical categories of CNs and arbitrary general terms can be

explained in terms of the ontological categories of *sortal* and *characterizing types* (Strawson 1959), which are roughly their ontological counterparts. Whilst the latter supply only a *principle of application* for the individuals they collect, the former supply both a principle of application and a *principle of identity*. A principle of application is one in accordance with which we judge whether a general term applies to a particular (e.g., whether something is a person, a dog, a chair or a student). A principle of identity supports the judgment whether two particulars are the same, i.e., in which circumstances the identity relation holds.

Cognitive psychologist John Macnamara (1986, 1994) has investigated the role of sortal concepts in cognition and provided a comprehensive theory for explaining the infant's process of learning proper and common nouns. He proposed the following example: suppose a little boy (Tom), who is about to learn the meaning of a proper name for his puppy. When presented with the word "Spot", Tom has to decide what it refers to. A demonstrative such as "that" will not suffice to determinate the bearer of the proper name. How to decide that the referent of "that", which changes all its perceptual properties, is still *Spot*? In other words, which changes can Spot suffer and still be the same? As Macnamara (among others) shows, answers to these questions are only possible if *Spot* is taken to be a proper name for an individual, which is an instance of a sortal universal. The principles of identity supplied by the sortals are essential to judge the validity of all identity statements. For example, if for an instance of the sortal *statue* losing one of its pieces will not alter the identity of the object, the same does not hold for an instance of *lump of clay*.

The claim that we can only make identity and quantification statements in relation to a sortal amounts to one of the best-supported hypothesis in the philosophy of language, namely, that the identity of an individual can only be traced in connection with a sortal universal, which provides a *principle of individuation* and *identity* to the particulars it collects (Macnamara 1986, 1994; Gupta 1980; Lowe 1989; van Leeuwen 1991).

As argued by Kripke (1980), a proper name is a rigid designator, i.e. it refers to the same individual in all possible situations, factual or counterfactual. For instance, it refers to the individual Mick Jagger both now (when he is the lead singer of Rolling Stones and 71 years old) and in the past (when he was the boy Mike Philip living in Kent, England). Moreover, it refers to the same individual in counterfactual situations such as the one in which he decided to continue attending the London School of Economics instead of pursuing a musical career. We would like to say that the boy Mike Philip is identical with the man Mick Jagger that he later became. However, as pointed out by Wiggins (2001) and Perry (1970), statements of identity only make sense if both referents are of the same type. Thus, we could not say that a certain boy is the same boy as a certain man since the latter is not a boy (and vice-versa). However, as Putnam put it, when a man x points to a boy in a picture and says "I am that boy," the pronoun "I" in question is typed not by man but by a type subsuming both man and boy (namely, person), which embraces x's entire existence (Putnam 1994). A generalization of this idea is thesis D, proposed by David Wiggins

(Wiggins 2001): *if an individual falls under two sortals F and F′ in the course of its history there must be exactly one ultimate sortal G that subsumes both F and F′*.

A proof of thesis D can be found in Guizzardi (2005). Intuitively, one can appreciate that it is not the case that two incompatible principles of identity could apply to the same individual *x*, otherwise *x* will not be a viable entity (determinate particular) (van Leeuwen 1991). For instance, suppose an individual *x* that is an instance of both *statue* and *lump of clay*. Now, the answer to the question whether losing one of its pieces will alter the identity of *x* is indeterminate, since each of the two principles of identity that *x* obeys imply a different answer. As a consequence, we can say that if two sortals F and F′ intersect (i.e., have common individuals in their extension), the principles of identity contained in them must be equivalent. Moreover, F and F′ cannot supply a principle of identity for *x*, since both sortals apply to x only contingently, and a principle of identity must be used to identify *x* in all possible worlds. Therefore, there must be a sortal G that supplies the principle of identity carried by F and F′. The unique ultimate sortal G that supplies the principle of identity for its instances is named a *substance sortal* or a *kind* (Gupta 1980; Guizzardi et al. 2004).

In the example above, person can only be the sortal that supports the proper name Mick Jagger in all possible situations because it applies necessarily to the individual referred to by the proper name, i.e., instances of person cannot cease to be so without ceasing to exist. As a consequence, the extension of a kind is world invariant, i.e., for all x, if x is an instance of a rigid type G then x must be an instance of G in all possible worlds. This meta-property of universals is called *modal constancy* (Gupta 1980) or *rigidity* (Guarino and Welty 2009). Every *kind* G is a rigid universal. Moreover, a kind G can be specialized into other sortals $F_1, \ldots, F_n$ that are themselves rigid. Take for instance the kind person. This kind can be specialized by the sortals male person and female person, which (in the biological sense) are themselves rigid sortals. I name the rigid sortals F_i that specialize a kind (thus inheriting its principle of identity) *subkinds* (Guizzardi et al. 2004).

Examples of *non-rigid* sortals include universals such as *boy* and *adult man* in the example previously discussed, but also *student*, *employee*, *caterpillar* and *butterfly*, *philosopher*, *writer*, *alive* and *deceased*. Actually, these examples of sortals are not only non-rigid, but they are *anti-rigid*. Non-rigidity is the simple logical negation of rigidity, i.e., a type is non-rigid if it does not apply necessarily to at least one of its instances. In contrast, a type is anti-rigid if it does not apply necessarily to all its instances. In other words, if a type F is anti-rigid then for all instances x of F there is a possible world in which x is not an instance of F. Sortals that possibly apply to an individual only during a certain phase of its existence are called *phased-sortals* (Wiggins 2001). As a consequence of thesis D, we have that: *for every phased-sortal PS that applies to an individual, there is a kind (substance sortal) S of which PS is a specialization*.

Although Frege argued at length that *"one cannot count without knowing what to count"* (Frege 1980), in artificial logical languages inspired by him, natural language general terms such as common nouns, adjectives and verbs are treated uniformly as predicates. For instance, if we want to represent the sentence "there

are tall men," in the Fregean approach of classical logic we would write $\exists x\,(man(x) \wedge tall(x))$. This reading puts the count noun man (which denotes a sortal) on an equal logical footing with the predicate tall. Moreover, in this formula, the variable x is interpreted as an alleged universal kind *Thing* (or entity). So, the natural language reading of the formula should be "there are things that have the property of being a man and the property of being tall." As argued in Hirsch (1982), concepts such as thing, (entity, element, among others) are *dispersive*, i.e., they cover many concepts with different principles of identity and do not denote sortals. This view is corroborated by many empirical studies in cognitive science (Xu et al. 2004).

The claims presented in this section are represented in a list of psychological claims proposed by Macnamara (1994) and are supported by a number of empirical studies (Xu et al. 2004; Bonatti et al. 2002; Waxman and Markow 1995; Booth and Waxman 2003). For instance, results from Xu et al. (2004) show that between 9 and 12 months of age a sortal-based system of individuation and identity emerges in infants' cognition. As remarked by the authors: "during this period, infants' worldview undergoes fundamental changes: They begin with a world populated with objects . . . By the end of the first year of life, they begin to conceptualize a world populated with sortal-kinds . . . In this new world, objects are thought of not as 'qua object' but rather 'qua dog' or 'qua table'."

9.3 A Logical System with Sortal and Characterizing Types

The formal characterization of the ontological distinctions discussed in the previous section requires some sort of modal treatment. In classical (extensional) modal logics, no distinction is made between different types of types. Types are represented as predicates in the language that divide the world (at each situation) into two classes of elements: those that fall under them and those that do not. This principle determines the extension of each type at each situation. Classical (one-place) predicates, being functions from worlds to sets of individuals, properly represent the *principles of application* that are carried by all types but fail to represent the *principles of identity*, which are unique of sortals. Equivalently, they treat all objects as obeying the same principle of identity.

Suppose that there is an individual person referred to by the proper name *John*. As discussed in the previous section, proper names for objects refer rigidly and, hence, if we say that John weights 80 kg at t_1 but 68 kg at t_2 we are in both cases referring to the same individual, namely the particular John. Now, let x_1 and x_2 be snapshots representing the projection of John at time boundaries t_1 and t_2, respectively. The truth of the statements *overweight(John,*t_1*)* and *overweight(John,*t_2*)* depends only on whether overweight applies to the states x_1 or x_2, respectively. In other words, the judgment if an individual i is an instance of a characterizing type G (e.g., overweight) in world w depends only whether the principle of application carried by G applies to the state of i in w. Now, how can one determine that, despite of possibly significant dissimilarities, x_1 and x_2 are states of the same particular John?

As previously argued this is done via a principle of cross-world identity and supplied by the substance sortal person, of which John is an instance.

These differences between sortals and characterizing types are made explicit in the formal language L_{sortal} defined in this section in the following way:

- The intension of the proper name John is represented by an *individual concept J*, i.e., a function that maps to a snapshop x_i of John in each possible world w. The notion of individual concepts, first introduced by Leibniz, refers to a singleton property that only holds for one individual;
- Sortal universals, such as person, are represented as *intensional properties*, which are functions from possible worlds to sets of individual concepts. For instance, for the sortal person there is a function ℓ that maps every world w to a set of individual concepts (including J). An individual x is a person in world w iff there is an individual concept $k \in \ell(w)$ such that $k(w) = x$;
- Individual concepts represent the principle of identity supplied by the universal person such that if $J(w) = x_1$ and $J(w') = x_2$ then we say that x_1 in w is the same person as x_2 in w', or in general: for all individuals x,y representing snapshots of an individual C of type T we say that x in w is the same T as y in w' iff C is in the extension of T and $C(w) = x$ and $C(w') = y$;
- Whilst the principle of identity is represented by sortal determined individual concepts that trace individuals from world to world, the principle of application considers individuals only at a specific world. For instance, John is overweight in word w iff *overweight(J(w),w)* is true.

Due to these considerations, in the language L_{sortal} presented below, the primitive elements in the domains of quantification are momentary states of objects, not the objects themselves. Ordinary objects of everyday experience (endurants, continuants) are instead represented by individual concepts. In the sequel, I formally define the syntax and semantics of L_{sortal}.

9.3.1 Syntax of L_{sortal}

Let L_{sortal} be a language of modal logics with identity with a vocabulary $V = (K,B,A,P,T)$ where: (a) T is a set of individual constants; (b) P is a non-empty set of n-ary predicates; (c) A is a set of phased-sortals (anti-rigid sortal types); (d) B is a set of subkinds; (e) K is a non-empty set of kinds (*substance sortal type*); (f) $R = K \cup B$ is named the set of rigid sortal types and the set $C = R \cup A$, the set of sortal types. The alphabet of L_{sortal} contains the traditional operators: $=$ (equality), $\neg$ (negation), $\rightarrow$ (implication), $\forall$(universal quantification), $\Box$ (necessity). The notions of term, sortal and formula are defined as follows:

Definition 1

1. All individual constants and variables are terms;
2. All sortal types belong to the category of sortal types;

3. If s and t are terms, then s = t is an atomic formula;
4. If P is a n-place predicate and $t_1 \ldots t_n$ are terms, then $P(t_1,..,t_n)$ is an atomic formula;
5. If A and B are formulas, then so are $\neg A$, $\Box A$, $(A \rightarrow B)$;
6. If S is a sortal classifier, x is a variable and A is a formula, then $(\forall S,x)A$ is a formula. ■

The symbols $\exists$ (existential quantification), $\wedge$ (conjunction), $\vee$ (disjunction), $\diamond$ (possibility) and $\leftrightarrow$ are defined as usual:

Definition 2

7. $(A \wedge B) =_{def} \neg (A \rightarrow \neg B)$;
8. $(A \vee B) =_{def} ((A \rightarrow B) \rightarrow B)$;
9. $(A \leftrightarrow B) =_{def} (A \rightarrow B) \wedge (B \rightarrow A)$;
10. $\diamond A =_{def} \neg \Box \neg A$
11. $((\exists S, x)\ A =_{def} \neg (\forall S, x)\ \neg A$
12. $((\exists !S, x)\ A =_{def} (\exists S, y)\ (\forall S, x)\ (A \leftrightarrow (x = y)$ ■

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

In L_{sortal}, sortal classifiers are never used in a predicative position. Therefore, if $S \in C$ is a sortal type, the predicate *s(x)* (in lowercase) is a meta-linguistic abbreviation according to the following definition.

Definition 3

$s\ (t) =_{def} (\exists S, x)\ (x = t)$ ■

According to this definition, the sentence "John is a man" is better rendered as "John is identical to a man". In opposition, in the sentence "John is tall," the copula represents the "is" of predication, which denotes a relation of mere equivalence.

9.3.2 *Semantics of L_{sortal}*

Definition 4 (Model Structure) A model structure for L_{sortal} is defined as an ordered couple $\langle W, D\rangle$ where: (i) W is a non-empty set of possible worlds; (ii) L_{sortal} adopts a varying domain frame (Fitting and Mendelsohn 1998) and, thus, instead of a set, D is a function that assigns to each member of W a non-empty set of elements. In order to avoid issues that are not germane to the purposes of this article, we simply assume here a universal accessibility relation between worlds (ibid.). ■

Given a model structure $M (= \langle W, D\rangle)$, the intension of an individual constant can be represented by an *individual concept*, i.e., a function i that assigns to each world $w \in W$, an individual in $D(w)$. Formally, we have that:

Definition 5 (Individual Concept) Let $M = \langle W, D\rangle$ and $U = \bigcup_{w \in W} D(w)$. An individual concept i in M is function from W into U, such that $i(w) \in D(w)$ in all worlds. For a given model structure M, we define I as a set of individual concepts defined for that structure. ■

The intension of an n-place predicate is defined (as usual) as an n-ary property, i.e., a function that assigns to each world $w \in W$ a set of n-tuples. If a tuple $\langle d_1 \ldots d_n\rangle$ belongs to the representation of a predicate at world w, then $d_1 \ldots d_n$ stand in w in the relation expressed by the predicate.

Definition 6 (Property) An n-ary property $(n > 0)$ in M is a function P from W into $\wp(D(w))^n$, i.e., if $\langle d_1 \ldots d_n\rangle \in P(w)$, then $d_1 \ldots d_n \in D(w)$. ■

The intension of sortal classifiers is defined such that both the principles of application and identity are represented. This is done by what Gupta (1980) calls *sorts*, i.e., *separated intensional properties*.

Definition 7 (Sort) Let $M = \langle W, D\rangle$ be a model structure. An intensional property in M is a function ℓ from W into the powerset of individual concepts in M (i.e., $\wp(I)$).

An intensional property assigns to each world a set of individual concepts, and it can be used to represent the intension of a sortal type in the following way. Suppose that ℓ represents the intension of the sortal type S and that the individual concept i belongs to ℓ at world w, i.e., $i \in \ell(w)$. Then i(w) is a S in w, and $i(w')$ is identical to i(w) in w.

Let ℓ be an intensional property in M, and let $L = \bigcup_{w \in W} \ell(w)$.

Now, let i,j be two individual concepts such that $i,j \in L$. We say that the intensional property ℓ is separated iff: if there is a world $w \in W$ such that $i(w) = j(w)$ then, for all $w' \in W$, $i(w') = j(w')$, i.e., $i = j$.

Finally, a sort in a model structure M is an intensional property that is separated. ■

The requirement of separation proposed in Gupta (1980) states, for example, that if two individual concepts for person, say 007 and James Bond, apply to the same object in a world *w* then they apply necessarily to the same object. This prevents unlawful conceptualizations in which a substantial individual splits or in which two individuals can become one while maintaining the same identity.

Given a sort ℓ in *M*, we designate by $\ell[w]$ the set of objects that fall under ℓ in *w*. Formally,

Definition 8 $\ell[w] = \{d: d \in D(w)$ and there is an individual concept $i \in \ell(w)$ such that $i(w) = d\}$. ■

Moreover, we define the set of objects in *w* that are *possibly* ℓ, i.e.,

Definition 9 $\ell\ |[w]| = \{d: d \in D(w)$ and there is an individual concept $i \in \ell(w')$ such that $i(w') = d\}$. ■

We now are able to define the notion of *counterpart* relative to a sort ℓ.

Definition 10 (Counterpart) We say that d in world w is the same ℓ as d' in w' iff there is an individual concept i that belongs to ℓ at some world (i.e., there is a w'' such that $i \in \ell(w''))$ and $i(w) = d$ and $i(w') = d'$. The ℓ counterpart in w' of the individual d in w is the unique individual d' such that d' in world w' is the same ℓ as d in w. ■

Finally, we are then able to define a model for L_{sortal}:

Definition 11 (Model) A model in L_{sortal} can be defined as a triple $\langle W, D, \delta\rangle$ such that:

1. 1.$\langle W, D\rangle$ is a model structure for L_{sortal};
2. δ is an interpretation function assigning values to the non-logical constants of the language such that: it assigns an individual concept to each individual constant c $\in$ T of L_{sortal}; an n-ary property to each n-place predicate p $\in$ P of L_{sortal}; a sort to each sortal type S $\in$ C of L_{sortal}.

   The interpretation function δ must also satisfy the following constraints:
3. If S $\in$ R then the sort ℓ assigned to S by δ must be such that: for all w, $w' \in W$, $\ell(w) = \ell(w')$, i.e., all rigid sortals are world invariant (modally constant);
4. Let S $\in$ (B $\cup$ A) be a subkind or an anti-rigid sortal type. Then, there is a kind S' $\in$ K such that, for all w $\in$ W, $\delta(S)(w) \subseteq \delta(S')(w)$;
5. Let S, $S' \in$ K be two kinds and let ℓ and ℓ' be the two sorts assigned to S and S' by δ, respectively. Then we have that: there is a w $\in$ W such that $\ell(w) \cap \ell'(w) \neq \varnothing$ iff $\ell = \ell'$, i.e., sorts representing kinds do not intersect unless they are identical. In other words, this restriction states that individuals belong to one single substance sortal, i.e., they obey one single principle of identity;
6. Let S $\in$ A be a phased (anti-rigid) sortal type. The sort ℓ assigned to S by δ must be such that: for all w $\in$ W, and for all individual concepts $i \in \ell(w)$, there is a world $w' \in W$ such that $i \notin \ell(w')$;

7. Let S, S′∈ K be two kinds and let ℓ and ℓ' be the two sorts assigned to S and S′ by δ, respectively. Then we have that: there is a w ∈ W such that $\ell[w] \cap \ell'[w] \neq \varnothing$ iff $\ell = \ell'$. Differently from (5) above, this restriction has it that individual states of objects can only be referred to by individual concepts of the same kind. ■

We are now able to define an assignment for L_{sortal}:

Definition 12 (Assignment) An assignment for L_{sortal} relative to a model $\langle W, D, \delta\rangle$ is a function that assigns to each variable of L_{sortal} an ordered pair $\langle \ell, d \rangle$, where ℓ is a sort relative to the modal structure $\langle W,D\rangle$ and $d \in U = \bigcup_{w \in W} D(w)$.

If a is an L_{sortal} assignment then $a_o(x)$ is the object assigned to variable x by a and $a_S(x)$ is the sortal to which x is bound. Moreover, it is always the case that $a_o(x) \in a_S(x)|[w]|$ for all variables. ■

Definition 13 An assignment a′ for L_{sortal} is an ℓ variant of a at x in w iff: a′ is just like a except perhaps at x (abbreviated as $a' \sim_x a$),

1. $a'_s(x) = \ell$,
2. $a'_o(x) \in \ell[w]$. ■

Definition 14 The w′ variant of an assignment a relative to w (abbreviated as f(w′,a,w)) is the unique assignment a′ that meets the following conditions:

(i) $a'_S(x) = a_S(x)$ at all variables x,
(ii) $a'_o(x)$ in w′ is the $a_S(x)$ counterpart of $a_o(x)$ in w relative,

at all variables x. ■

Definition 15 (Truth-Theoretical Semantics) Finally, let α be an expression in L_{sortal}, and let the semantic value of α at world w in model M relative to assignment *a* be the value of the valuation function $\nu^w_{M,a}$.

With these definitions, we can define the semantics of L_{sortal} as follows:

(a) If α is an individual constant or a sortal type, then $\nu^w_{M,a}(\alpha) = \delta(\alpha)(w)$.
(b) If α is variable, then $\nu^w_{M,a}(\alpha) = a_o(\alpha)$
(c) If α is an atomic formula $t_1 = t_2$, then $\nu^w_{M,a}(\alpha) = T$ if $\nu^w_{M,a}(t_1) = \nu^w_{M,a}(t_2)$. Otherwise $\nu^w_{M,a}(\alpha) = F$.
(d) If α is an atomic formula $P(t_1 \ldots t_n)$, then $\nu^w_{M,a}(\alpha) = T$ if $\langle \nu^w_{M,a}(t_1) \ldots \nu^w_{M,a}(t_n)\rangle \in \delta(P)(w)$. Otherwise $\nu^w_{M,a}(\alpha) = F$.
(e) If α is the formula ¬A, then $\nu^w_{M,a}(\alpha) = T$ if $\nu^w_{M,a}(A) = F$. Otherwise $\nu^w_{M,a}(\alpha) = F$.
(f) If α is the formula $(A \rightarrow B)$, then $\nu^w_{M,a}(\alpha) = T$ if $\nu^w_{M,a}(A) = F$ or $\nu^w_{M,a}(B) = T$. Otherwise $\nu^w_{M,a}(\alpha) = F$.
(g) If α is the formula (∀S,x)A, then $\nu^w_{M,a}(\alpha) = T$ if $\nu^w_{M,a'}(A) = T$ for all assignments a′ which are δ(S) variants of a at x in w. Otherwise $\nu^w_{M,a}(\alpha) = F$.
(h) If α is the formula □A, then $\nu^w_{M,a}(\alpha) = T$ if $\nu^{w'}_{M,f(w',a,w)}(A) = T$ for all w′∈ W. Otherwise $\nu^w_{M,a}(\alpha) = F$. ■

9.3.3 Discussion

The language L_{sortal} has been proposed based on the first of four systems introduced by Anil Gupta in his *Logic of Common Nouns* (Gupta 1980). Gupta, however, does not elaborate on different types of sortals. Consequently, restrictions (3) to (7) on δ in definition 11 are simply not defined in his system. Restriction (7), in particular, would have to be rejected by Gupta, as a consequence of his contingent (or relative) view of identity. Note that restriction (7) implies (5) but not vice-versa.

It is widely accepted that any relation of identity must comply with Leibniz's law: if two individuals are identical then they are necessarily identical (van Leeuwen 1991). Relativists, however, adopt the thesis that it is possible for two individuals to be identical in one circumstance but different in another. A familiar example, cited by Gupta, is that of a statue and a lump of clay. The argument proceeds as follows: Suppose that in world w we have a statue *st* of the Dalai Lama which is identical to the lump of clay *loc* that this statue is made of. In w, *st* and *loc* have exactly the same properties (e.g., same shape, weight, color, temperature, etc.). Suppose now that in world w', a piece (e.g., the hand) is subtracted from *st*. If the subtracted piece is an inessential part of a statue then the statue st' that we have in w' is identical to *st*. In contrast, the lump of clay *loc'* which st' is made of is different from *loc*. In summary, we have in w' the same statue as in w but a different lump of clay. In Gupta's system, without restriction (7), we have it that for two individual concepts i and j such that $i(w) = j(w)$, it remains possible a world w' such that $i(w') \neq j(w')$. In other words, the formula (α) $(\exists \text{Statue,x } (\text{x} = \text{dl}) \wedge \exists \text{LoC,y } (\text{y} = \text{dl}) \wedge \Diamond(x \neq y))$ is satisfiable.

I reject this line of reasoning for two reasons. Firstly, I support the view that Leibniz's rule must hold for a relation to be considered a relation of identity, otherwise, any equivalence relation such as *being an instance of the same class* would have to be considered a relation of identity. Secondly, if Gupta's primitive elements are thought of as momentary states, then (α) does not in fact qualify as a statement of relative identity. It actually expresses that two objects can coincide (i.e., share the same state) in a world w but not in a different world w' (van Leeuwen 1991). Notice, however, that if restriction (7) is assumed, formula (α) is no longer satisfiable.

Proof (a) if (x = dl) is true then there is an individual concept st of statue that refers in the actual world w to the same entity d as dl; (b) if (y = dl) is true then there is an individual concept loc of LoC that refers in the actual world w to the same entity d as dl; (c) by transitivity of equality, st and loc refer to the same d in world w and, consequently, d is then both of the kind Statue and of the kind LoC in w; (e) due to (7), the intensions of Statue and LoC are identical; (f) finally, due to separation, st and loc must coincide in every world. □

Now we are in a position to choose between two alternatives related to the interpretation of momentary states. The first is to assert restriction (7) and take a *multiplicationist* (Guizzardi 2005) stance such that *st* and *loc* do not actually share

the same state in *w* in the strong sense. Rather, I consider the states *st(w)* and *loc(w)* to be numerically different albeit instantiating the same types (properties).

A second stance is to assume that two continuants (endurants) can indeed share a state in the numerical sense. If we accept this, a simple way of modifying L_{sortal} to account for coincidence as manifested in Gupta's system consists in:

(a) removing the constraint (7) in definition 11;
(b) including the operator $\approx$ for coincidence, with the following semantics: If α is an atomic formula $t_1 \approx t_2$, then $\boldsymbol{\nu}^{w}_{M,a}(\alpha) = T$ if $\boldsymbol{\nu}^{w}_{M,a}(t_1) = \boldsymbol{\nu}^{w}_{M,a}(t_2)$. Otherwise $\boldsymbol{\nu}^{w}_{M,a}(\alpha) = F$;
(c) defining the identity relation between individual constants as $(t_1 = t_2) =_{def} \Box(t_1 \approx t_2)$, i.e., two continuants are identical if they coincide in every possible world.

A version of L_{sortal} which takes the *multiplicationist* stance can serve in support of two goals: defining the semantics of object-oriented and database languages in computer science, and to circumvent some of the limitations in representing modal (temporal) information in terminological languages such as OWL (Web Ontology Language).[1] In the sequel, I will briefly present an example of the first goal. For an example of the latter, I refer readers to Zamborlini and Guizzardi (2010). In the next section, I employ the proposed framework to address issues of cross-world identity and dynamic classification in conceptual spaces.

For instance, in the Unified Modeling Language (UML),[2] a *de facto* standard for conceptual modeling in computer science, types are represented in so-called class diagrams.[3] In contrast, the instances of these types are represented in object diagrams. See Figs. 9.1 and 9.2 below. In Fig. 9.1, we have a representation of the type person characterized by the properties *name*, *social security number*, *age* and *height*, as well as the type car characterized by the properties *chassis number*, *color*, *kilometer count* and *manufacturing date*. Moreover, the diagram represents a relational property *owns*, defined between instances of person and instances of car together with some integrity constraints on this relational property (while people can own *zero-to-many* cars, we assume that a car must be owned by exactly one person). In Fig. 9.2, we have a representation of an instance of the type person (John) and two instances of the type car (car_1 and car_2), as well as representation of two instances of the relational property *own*. Notice that individuals that appear in an UML object diagram (such as is given in Fig. 9.2) are not endurants. They are not persons like you and me, or cars like mine or yours. These are snapshot entities, momentary states of endurants. However, the instances of a UML class diagram (Fig. 9.1) are not snapshot entities; instead they are so-called *oid (object identifiers)*. Although

[1] www.w3.org/2004/OWL/

[2] http://www.uml.org/

[3] What are termed classes in UML are akin to what I name types here, not to the well-known set-theoretical notion of classes. In other words, classes in UML are intensional not extensional entities.

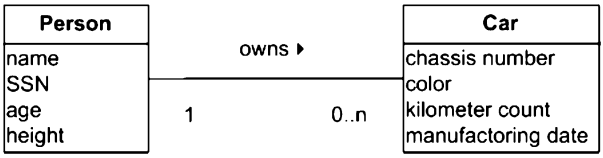

Fig. 9.1 Representation of a Conceptual Schema at the type level in the UML modeling language in the so-called class diagrams

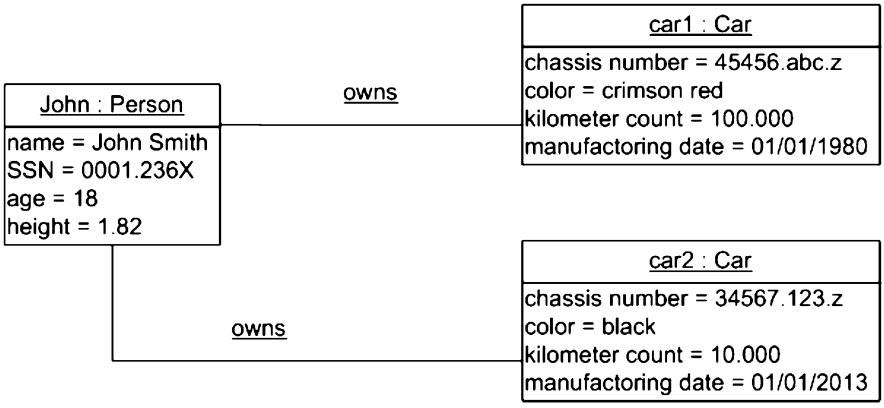

Fig. 9.2 Representation of a Conceptual Schema at the instance level in the UML modeling language in the so-called instance diagrams

this is not made explicit in the definition of the UML standard, an *oid* such as *John* (or car_1 and car_2) is supposed to connect the various snapshot entities (representing momentary states of John) that appear in different UML instance diagrams (hence, the identifier John of type person – symbolized as John:Person - in the header in Fig. 9.2). In summary, *oids* can be interpreted in L_{sortal} as individual concepts; entities in an object diagram can be interpreted instead as momentary states of objects in the sense discussed in Sects. 9.2 and 9.3. It is important to highlight that, in UML, snapshot entities are connected to exactly one *oid*. So, even if two snapshot entities in an instance diagram have the exact same value for all its properties, they still represent two numerically different individuals. Finally, although UML does not make a distinction between sortals and characterizing types, this distinction is available in an evolution of UML for the purpose of conceptual modeling called OntoUML (Guizzardi 2005). In OntoUML, *oids* are defined by classes representing kinds (substance sortals) in the model.

9.4 Cross-World Identity and Classification in Conceptual Spaces

9.4.1 Conceptual Spaces

A proposal to model the relation between the properties and concepts (types) classifying an individual and their representation in human cognitive structures is presented in the theory of *conceptual spaces* developed by the Swedish philosopher and cognitive scientist Peter Gärdenfors (Gärdenfors 2000). The theory is based on the notion of *quality dimension*. The idea is that several perceivable or conceivable properties are associated to quality dimension in human cognition. For example, height and mass are associated with one-dimensional structures featuring a zero point (i.e., isomorphic to the half-line of nonnegative numbers). Other properties such as color and taste are represented by several dimensions. For instance, taste can be represented as a tetrahedron space comprising the dimensions saline, sweet, bitter and sour, and color can be represented in terms of the dimensions hue (a polar dimension), saturation and brightness (two linear dimensions). An illustration of a *color domain* is depicted in Fig. 9.3 below.

According to Gärdenfors, some quality dimensions (especially those related to perceptual qualities) seem to be innate or developed very early in life. For instance, the sensory moments of color and pitch are strongly connected with the neurophysiology of their perception. Other dimensions are introduced by science or human conventions. For example, the representation of Newton's distinction between mass and weight is not given by the senses but has to be learned by adopting the conceptual space of Newtonian mechanics.

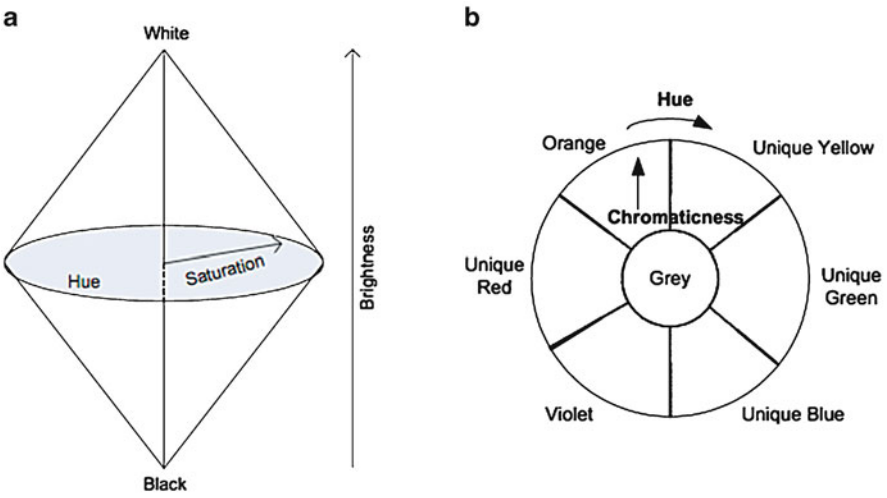

Fig. 9.3 Representations of a color spindle (quality domain for color)

Zenker and Gärdenfors (2015), in this volume, distinguish between *integral* and *separable* quality dimensions: "Dimensions are said to be integral if, to describe an object fully, one cannot assign it a value on one dimension without giving a value on the other. For example, an object cannot be given a hue without giving it a brightness value. Or the pitch of a sound always goes along with its loudness . . . Dimensions that are not integral are said to be separable, as for example the size and hue dimensions". They then define a *quality domain* as "a set of integral dimensions that are separable from all other dimensions". Finally, a *conceptual space* is defined as "collection of one or more domains" (Gärdenfors 2000, p. 26).

Gärdenfors emphasizes that the notion of *conceptual space* should be understood literally, i.e., quality dimensions, quality domains and conceptual spaces are endowed with certain geometrical structures (topological or ordering structures) that constrain the relations between its constituting dimensions. In particular, Gärdenfors uses the notion of a convex region in a metric space to define what he calls a *quality region*. For instance, the different regions in the color circle of Fig. 9.3b define quality regions in that domain. According to him, only attributes representing genuinely substantial properties will form quality regions in a conceptual space. This allows for a geometrical grounding of the difference between what David Lewis (1986) called *natural attributions*, as opposed to *abundant attributions*. In the conceptual space model, natural attributions (e.g., red, person, car) will form convex regions, but abundant attributions will not (e.g., *not-red*, *being-a-car-or-an-apple*).

Finally, Gärdenfors makes the following distinction between what he calls *concepts* and *properties* (Gärdenfors 2000): "Properties . . . form as special case of concepts. I define this distinction by saying that a property is based on single domain, while a concept may be based on several domains." In other words, properties define regions completely contained in a quality domain while concepts define regions that cross over multiple quality domains.

9.4.2 Individuation, Identity and Dynamic Classification in Conceptual Space

Regarding the notions of principle of application and principle of identity discussed throughout this article, a number of remarks can be made regarding the conceptual spaces model.

9.4.2.1 Principles of Application in Conceptual Spaces

In the conceptual space model, an individual is identified by a point in a conceptual space. An object (continuant, endurant) like you and me, my car, your house, Susan's cat, the planet Mars, and the Monalisa are identified by a vector in a multi-dimensional space (a hyperspace) so that each component (coordinate) of

a vector represents a value of a property on a given quality dimension (e.g., my height, the color of my car, the price of my house). In fact, Gärdenfors admits to the Leibnizian principle of identity for all individuals, i.e., two individuals are the same (in a numerical sense) iff they are represented by the same point in a conceptual space. Provided that individuals are points in a conceptual space, the *principle of application* of a given type can be represented by the geometrical notion of spatial containment in a given region. For instance, we know that my car is red because the coordinates that represent the color of my car (vector) lie within the red region in the color space.

9.4.2.2 Comparing the Sortal and Characterizing Types and the Concepts vs. Properties Distinctions

Gärdenfors' distinction between properties and concepts does not correspond to that between sortal and characterizing types discussed in Sect. 9.2. For once, there are characterizing types that will correspond to regions crossing multiple quality domains. Examples include the types *physical object* (as a supertype of houses, cars, persons), *insurable items* (as a supertype of persons, houses, cars, works of art, buildings). Moreover, the points (individuals) in a region defined by a property (e.g., red) in the sense of Gärdenfors are exemplars not of objects but of what is termed a *quality value* (i.e., the super-determinate value of a quality, a trope, a property instance, an abstract particular) (Guizzardi 2005). If instead, one is willing to conceptualize the object type red (whose instances would include a red car, a red flag, a red apple, a red building), then the corresponding region crosses multiple quality domains (e.g., my red car will be fully conceptualized on many dimensions that are separable). We conclude that, in its present state, the theory of conceptual spaces does not make the distinction between sortal and characterizing types, i.e., a distinction between types that merely offer a principle of application to its instances and types that also offer a principle of identity. As previously mentioned, the theory can adequately represent the former (e.g., my car is an instance of the object type red if it lies within the cross-domain Red region) but not the latter. Given the purposes of this paper, from now on, I focus on regions of conceptual spaces associated to object types. I assume that ordinary objects (in the sense investigated here) will always be associated to *concept regions*, i.e., to regions crossing multiple domains.

9.4.2.3 Limitations of Conceptual Spaces Regarding Cross-World Identity and Dynamic Classification

A region in a conceptual space representing a type such as dog must represent not only the current dogs that exist now but, as put by Gauker (2007), they must *"comprise all and only dogs, since the concept dog correctly applies to each and every dog (that ever has been or ever will be) and to nothing else."* However, as Gauker notes, the regions defined in a conceptual space are static (fixed). In fact,

if an individual is represented by a point in a conceptual space, and one adopts a Leibnizian principle of identity (as Gärdenfors does), then an entity cannot suffer any change without ceasing to be the same. As a consequence, as a representation for types (concepts), regions in a conceptual space seem to be only able to represent *rigid types* and immutable individuals. Gauker makes a similar point arguing that *similarity spaces theories* of concepts, in general, and the theory of conceptual spaces in particular, cannot support the structure of judgments (Gauker 2007).

To illustrate this point, suppose a situation in which an individual, John, is a college student. According to the conceptual space theory, this is represented by having a point *x* represent John in the college student region of a conceptual space. Now, suppose that John ceases to be a college student. John must now be represented by a new point *y* outside the college student region in that conceptual space. Notice that by definition the two points x and y (the two vectors containing different component values) are different. Quoting Gauker, the following question arises: *"in what sense the earlier point in college student region represents the same thing as the later point outside the college student region?"* How can we say that these two points represent the same object?

Gauker's example in fact is slightly different. He exposes a situation in which someone initially *judges* John to be a college student (so her belief that John is a college student is represented by a point in the college student region representing John) but later learns that John is in fact not one (so her belief that John is not a college student is represented by a point outside the college student region representing John). Although Gauker's example pertains to belief revision (learning about individuals), for the sake of my argument, its point is exactly the same. After all, someone's mistaken belief about John can be thought of as a conception of John in a counterfactual situation, i.e., one in which the very same individual has *some* properties different from those he now has (like the counterfactual situation where Mick Jagger never quits the London School of Economics and never leads the Rolling Stones). Furthermore, in order to learn things about John, one must recognize or conceptualize the same individual in different (counterfactual) situations as the very same individual. After all, it is not the case that all properties of an individual are manifested in each of encounters with them (Macnamara 1986).

9.4.2.4 Kind-Dependent Identity

In our working example, merely including an extra time dimension to the space of persons (and, hence, to that of college students) will not suffice to address the above issue. Suppose we were to add an extra coordinate to all vectors representing individuals in the person space (and all its sub-regions including that of college student). John being a college student at time t_1 would then be represented by the point x ($\langle x_1 \ldots x_n, t_1 \rangle$), having values for a number of coordinates (including those referring to properties of students); John not being a student at time t_2 would then be represented by the point y ($\langle y_1 \ldots y_n, t_2 \rangle$), having values for a number of coordinates *but not those referring to properties of students*. Notice that our original

question still persists: how can we judge that x and y are the same individual in two different situations? Summing up all these different points (among possibly others) and deciding that John actually represents a sequence of time-indexed vectors does not offer any explanatory power. After all, the problem is exactly one of deciding which points should be part of this sequence of vectors. In other words, what kind of changes can an individual suffer and still be the same individual! There must be something that remains the same in all points representing the same individual. Or, using the terminology of similarity spaces, there must be a set of non-zero values for all points representing John (perhaps some of these values are even immutable across these points). *The specific set of these values depends on what kind of entity is being represented by these points*, i.e., it is because John is a person that all points representing John must have values representing properties that must be present for instances of the concept person, regardless if he is a college student in a particular situation, or not.

As defended here, to decide which points constitute the sequence of points representing an individual in a time-indexed conceptual space, we need the support of a kind K. This kind K will supply a principle of cross-world identity which reports on the properties that must be present in all instances of K (i.e., which dimensions must have non-zero values for points in a given region) and the property values that must remain the same for an entity to remain the same K (i.e., which coordinates must be present for points in a conceptual space to represent the same instance of K).

9.4.2.5 Relating L_{sortal} and Conceptual Spaces

As previously discussed, the conception of object types as regions in a conceptual space can adequately represent the *principle of application* of a given type. However, one should notice that the points in these regions should not be interpreted as objects (continuants, endurants) but as *momentary states of objects*, i.e., the sort of individuals pertaining to the domain of quantification of characterizing types as discussed in Sects. 9.2 and 9.3 (i.e., members of the set $U = \bigcup_{w \in W} D(w)$).

In other words, points in a cross-domain region of a conceptual space corresponding to an object concept should be interpreted as *qualitative characterizations of states of objects* falling under that concept. In particular, unary properties standing for characterizing types in L_{sortal} should correspond to cross-domain regions (object concept regions) in a conceptual space.

My example in the previous section is about time. However, to address objections such as Gauker's, we should take a more general view on an indexing dimension. In other words, the points on such a dimension and the structure of that dimension should correspond exactly to worlds and their accessibility relations, respectively, as discussed in Sect. 9.3. However, for the sake of maintaining generality over the possible interpretations of worlds, I will not assume here that world-structures are additional dimensions on conceptual spaces. Instead, they will be defined as part of an additional structure used for the representation of sortal concepts in the

sequel. An additional reason for not including world structures as dimensions in our conceptual spaces is the idea that points, which represent momentary states of entities independently of a world structure are sufficient for applying a principle of application.

I hold that enduring objects of everyday experience cannot be directly represented by standard conceptual spaces. In other words, the instances of sortal types like person, organization, country, car, president, child, planet or statue cannot be directly mapped to points in a conceptual space. To represent such sortal types, we must define associate structures that define suitable projections into conceptual spaces. These structures associate to each sortal type a *sort* ℓ (i.e., a *separated intensional property*) whose extension contains *individual concepts*. Individual concepts can be thought of as projections into a conceptual space defining a suitably constrained *set of points* that represent counterparts in different worlds of the same ordinary object. Sorts, in turn, are sets of individual concepts and, hence, can be thought as projections into conceptual spaces that define regions containing suitably constrained *sets of sets of points*, representing states of ordinary objects of the same sortal type.

So, whilst a point x in a conceptual space can be directly judged to be a *red individual,* an *electrically charged entity* or a *physical object*, that point can only be judged to be a state of person in world w if x belongs to $\ell[w]$. Moreover, whilst regions associated to characterizing types can be defined as similarity regions based uniquely on the similarity of basic points, regions associated to sortal types are *projected* into conceptual spaces by the principle of identity carried by that type. To put it in another way, the latter type of similarity regions are defined in terms of sets of points selected by individual concepts, not in terms of basic points.

In summary, instances of sortals types should be represented by *individual concepts* representing a principle of identity, supplied by the kind they instantiate, which can trace the identity of the same individual by referring to (qualitatively distinct) states in different worlds (represented by points in a conceptual space). These individual concepts are supplied by kinds (substance sortals). However, they can also be dynamically classified possibly under a number of anti-rigid types representing contingent (accidental) properties that can inhere in these individuals. So, returning to our working example, the same individual person, John, can fall in the extension of the type student in a number of situations (in which the states of John will be represented by points in the student region of the person space), and it can fall outside this extension in a number of other situations (in which the states of John will be represented by points outside the student region in the person space). Nonetheless, it is the very same individual, John, that maintains its numerical identity regardless of these contingent (de)classifications as a student.

Finally, given the non-multiplicationalist stance adopted here, the same point in a conceptual space can belong to regions associated to different object types (concepts), i.e., the same point can represent states of individuals falling under different concepts. Moreover, on this stance, two individuals that share a qualitatively indistinguishable state in a given world do have the same state in a numerical sense.

In other words, if the statue *st* and the lump of clay *loc* coincide in world *w* (i.e., $st(w) \approx loc(w)$) then they refer, in that world, to the very same point in a conceptual space.

9.5 Related Work

Modal notions such as were discussed in this paper have been employed by Guarino and Welty (2009) in a number of publications as a way to formally characterize the ontological distinctions comprising the OntoClean evaluation approach for taxonomic structures. OntoClean clearly distinguishes sortal and characterizing types according to their ontological status. However, in the formalizations of that approach, a classical system of modal logics is employed where the focus is on distinguishing between properties w.r.t. to their modal meta-properties (e.g., rigidity versus non-rigidity). As a consequence, these formalizations fail to capture a fundamental distinction between sortal and characterizing types and the unique role of the former category in providing a principle for trans-world identity for objects.

The idea of representing objects of ordinary experience by individual concepts is similar to the solution adopted in the GFO foundational ontology (Heller et al. 2004) in which individual concepts for objects are called *abstract substances* or *persistents*. The notion of a momentary state of objects adopted here is similar to that of *presentials* there. As demonstrated by Heller and Herre, a language such like the one proposed here can play an important role in relating *endurantistic (3D)* and *perdurantistic (4D)* views of entities (i.e., views of entities as space-extended objects with those of entities as spatiotemporal processes). However, in contrast to our approach, GFO does not elaborate on different categories of types (*viz.* kinds, subkinds, phased-sortals and characterizing types). Consequently, no connection between types and identity is developed, and the approach makes no distinction between types that aggregate essential properties (rigid types) and those that aggregate merely contingent ones. Accounting for such distinctions is fundamental not only from a theoretical point of view but also for a number of applications in computer science (Guizzardi 2005). Furthermore, as discussed in Sect. 9.4, this distinction also plays an important role in addressing a criticism targeted at conceptual spaces by Gaulker (2007).

9.6 Summary

I presented a system of modal logic with sortal restricted quantification to suitably capture the intended semantics of a philosophically and cognitively well-founded theory of object types. The proposed logical system formally characterizes the distinction between sortals and general property types where the former exclusively supplies a principle of persistence and cross-world identity to its instances. As a

result, we can address the limitations of classical (unrestricted extensional) modal logics which reduce ontologically very different categories to the same logical footing, and advance proposals such as in Gupta (1980) by: (i) refining the notion of sortal types, considering the distinction between substance, rigid and phased-sortals; and (ii) proposing a system that avoids reducing the relation of identity to a mere relation of equivalence. Finally, I also showed how this proposal can complement the theory of conceptual spaces by offering an account for kind-supplied principles of cross-world identity. The proposal is in line with a number of empirical results in cognitive psychology and that can addresses an important criticism of the conceptual spaces model regarding object identity. As I demonstrate here, without addressing issues related to cross-world object identity, the conceptual spaces model is not properly equipped for serving as a general model for cognitive semantics, as it is not properly equipped for defining the semantics of linguistic entities as fundamental as proper names.

References

Bonatti, L., Frot, E., Zangl, R., & Mehler, J. (2002). The human first hypothesis: Identification of conspecifics and individuation of objects in the young infant. *Cognitive Psychology, 44*, 388–426.

Booth, A. E., & Waxman, S. R. (2003). Mapping words to the world in infancy: Infants' expectations for count nouns and adjectives. *Journal of Cognition and Development, 4*(3), 357–381.

Fitting, M., & Mendelsohn, R. L. (1998). *First-order modal logic, synthese library* (Vol. 277). Dordrecht: Kluwer Academic Publisher.

Frege, G. (1980). *The foundations of Arithmetic: A logic-mathematical enquiry into the concept of number*. Translated from the 1884 original by J. L. Austin, Northwestern University Press; 2nd Revised edition.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, USA: MIT Press.

Gauker, C. A. (2007, September). Critique of the similarity space theory of concepts. *Mind & Language*, *22*(4), 317–345.

Guarino, N., & Welty, C. (2009). An overview of OntoClean. In S. Staab & R. Studer (Eds.), *Handbook on ontologies* (2nd ed., pp. 201–220). Berlin: Springer.

Guizzardi, G. (2005). *Ontological foundations for structural conceptual models* (Telematics Instituut Fundamental Research Series, No. 015). ISSN 1388-1795, The Netherlands.

Guizzardi, G., Wagner, G., Guarino, N., & van Sinderen, M. (2004). An ontologically well-founded profile for UML conceptual models. In A. Persson & J. Stirna (Eds.), *Lecture notes in computer science* (pp. 112–116). Berlin: Springer. ISBN 3-540-22151-4.

Gupta, A. (1980). *The logic of common nouns: An investigation in quantified modal logic*. New Haven: Yale University Press.

Heller, B., Herre, H., Burek, P., Loebe, F., & Michalek, H. (2004). *General ontological language* (Technical Report no. 7/2004). Institute for Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Germany.

Hirsch, E. (1982). *The concept of identity*. New York/Oxford: Oxford University Press.

Kripke, S. (1980). *Naming and necessity*. Princeton: Wiley-Blackwell.

Lewis, D. K. (1986). *On the plurality of worlds*. Oxford: Blackwell.

Lowe, E. J. (1989). *Kinds of being: A study of individuation, identity and the logic of sortal terms*. Oxford: Blackwell.

Macnamara, J. (1986). *A border dispute, the place of logic in psychology*. Cambridge: MIT Press.
MacNamara, J. (1994). Logic and cognition. In J. MacNamara & G. Reyes (Eds.), *The logical foundations of cognition, Vancouver studies in cognitive science* (Vol. 4, pp. 11–34). New York: Oxford University Press.
Perry, J. (1970). The same F. *Philosophical Review, 79*(2), 181–200.
Putnam, H. (1994). Logic and psychology. In J. MacNamara & G. Reyes (Eds.), *The logical foundations of cognition, Vancouver studies in cognitive science* (Vol. 4, pp. 35–42). New York: Oxford University Press.
Strawson, P. F. (1959). *Individuals. An essay in descriptive metaphysics*. London/New York: Routledge.
van Leeuwen, J. (1991). *Individuals and sortal concepts: An essay in logical descriptive metaphysics*, PhD thesis, University of Amsterdam, Amsterdam.
Waxman, S. R., & Markow, D. R. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29*, 257–302.
Wiggins, D. (2001). *Sameness and substance renewed*. Cambridge, UK: Cambridge University Press.
Xu, F. (2004). Categories, kinds, and object individuation in infancy. In L. Gershkoff-Stowe & D. Rakison (Eds.), *Building object categories in developmental time* (pp. 63–89). Papers from the 32nd Carnegie Symposium on Cognition, New Jersey, Lawrence.
Zamborlini, V., & Guizzardi, G., (2010). *On the representation of temporally changing information in OWL*. IEEE 5th Joint International Workshop on Vocabularies, Ontologies and Rules for The Enterprise (VORTE) – Metamodels, Ontologies and Semantic Technologies (MOST), together with 15th International Enterprise Computing Conference (EDOC 2010) (pp. 283–292). Vitória, Brazil.
Zenker, F., & Gärdenfors, P. (2015). *Communication, rationality, and conceptual changes in scientific theories*. In this volume.

Chapter 10
A Cognitive Architecture for Music Perception Exploiting Conceptual Spaces

Antonio Chella

Abstract A cognitive architecture for a musical agent is presented. The architecture extends an architecture for computer vision previously developed by the author by taking into account many relationships between vision and music perception. The focus of the agent architecture is an intermediate conceptual area between the subconceptual and linguistic areas. A conceptual space for the perception of tones and intervals is thus presented, based on the dissonance measure of the tones. Problems and future works of the proposed approach are finally discussed.

10.1 Introduction

Gärdenfors (1988), in the seminal paper on "Semantics, Conceptual Spaces and Music," discusses a program for musical analysis based on conceptual spaces and inspired by the framework proposed by Marr (1982). In details, the first level of the program is related with pitch identification and it feeds input to the subsequent levels. The second level is related with the analysis of musical intervals; this level may also take into account the cultural background of the listener. The third level concerns tonality; here, scales are identified and the concepts of chromaticity and modulation arise. The fourth level is related with the interplay of pitch and time. In facts, time is processed by a hierarchy of levels related with temporal intervals, beats and rhythmic patterns. At the fourth and last level, eventually pitch and time merge together.

Bregman (1994) in his book on "Auditory Scene Analysis" discusses in great details how the Gestalt principles are at the basis of visual, music and speech perception. Indeed, many computational models of music perception take into account Gestalt principles, see Deutsch (2013) for a review.

Tanguiane (1993) considers three different levels of analysis distinguishing between statics and dynamics perception in vision and music. The first visual level in statics perception is the level of pixels; in the analogy of the image level of

A. Chella (✉)
Department of Chemical, Management, Computer, Mechanical Engineering, University of Palermo, Viale delle Scienze, building 6, 90128 Palermo, Italy
e-mail: antonio.chella@unipa.it

F. Zenker, P. Gärdenfors (eds.), *Applications of Conceptual Spaces*,
Synthese Library 359, DOI 10.1007/978-3-319-15021-5_10

Marr, this corresponds to the perception of frequencies. At the second level, the perception of simple patterns in vision corresponds to the perception of single notes. Finally, at the third level, the perception of structured patterns corresponds to the perception of chords. Concerning dynamic perception, the first level is the same as in the case of static perception, i.e., pixels and frequencies, while at the second level the perception of visual objects corresponds to the perception of notes. At the third level the perception of visual trajectories corresponds to the perception of music melodies. Tanguiane also proposes a hierarchy of levels related with time.

The ecological approach to perception introduced by Gibson (1979) inspired many works related with the *music as motion* theory, see Clarke (2005), Windsor and de Bèzenac (2012), Krueger (2009), Leman (2008), and London (2004), for interesting examples. According to this approach, the perception of music is related with motions and gestures induced by the music itself.

Several computational models of music cognition have been proposed in the literature; see Wiggins et al. (2009) and Temperley (2012) for reviews. Representative systems are, among others: MUSACT (Bharucha 1987, 1991) based on various kinds of neural networks; the IDyOM system based on a probabilistic model of music perception (Pearce and Wiggins 2004, 2006; Wiggins et al. 2009); the Melisma system developed by Temperley (2001) and based on preference rules of symbolic nature; the HARP system, aimed at integrating symbolic and subsymbolic levels (Camurri et al. 1994, 1995).

Concerning the relationships between music perception and conceptual spaces, Forth et al. (2010) discuss the relationships between the abstract notion of conceptual spaces in the sense of Boden (2004), and the concrete, geometrically grounded notion of conceptual spaces in the sense of Gärdenfors from the point of view of creative musical systems (Wiggins 2006). Bååth et al. (2013) discuss a model for rhythm perception based on conceptual spaces inspired by the cognitive model proposed by Desain and Honing (2003).

Here, a cognitive architecture for a musical agent is sketched. The cognitive architecture extends the architecture for computer vision previously developed by the author and aimed at the integration of symbolic and subsymbolic approaches. The architecture has been employed for static scenes analysis (Chella et al. 1997, 1998), dynamic scenes analysis (Chella et al. 2000), reasoning about robot actions (Chella et al. 2001), robot self recognition (Chella et al. 2003), and recently for modeling some aspects of robot self-consciousness (Chella et al. 2008). The presented extension of the architecture takes into account many of the above outlined relationships between vision and music perception.

In the following, Sect. 10.2 outlines the cognitive architecture for the musical agent while Sect. 10.3 describes the adopted conceptual space for the perception of tones and intervals. Section 10.4 outlines the linguistic area of the agent cognitive architecture, and finally, Sect. 10.5 discusses some of the problems related with the proposed approach and it outlines some future extension of the cognitive architecture.

10.2 The Cognitive Architecture

The proposed cognitive architecture for music perception is sketched in Fig. 10.1. The areas of the architecture are concurrent computational components working together on different commitments. There is no privileged direction in the flow of information among them: some computations are strictly bottom-up, with data flowing from the subconceptual up to the linguistic through the conceptual area; other computations combine top-down with bottom-up processing.

The *subconceptual* area is concerned with the processing of data coming from the sensors. Here, information is not yet organized in terms of conceptual structures and categories. Instead, in the *linguistic* area, representation and processing are based on a logic-oriented formalism.

The *conceptual* area is an intermediate level of representation between the subconceptual and the linguistic areas and it is based on conceptual spaces. Here, data is organized in conceptual structures that are independent of linguistic description. The symbolic formalism of the linguistic area is then interpreted on suitable aggregations of these structures. The conceptual space therefore acts as a sort of workspace in which low-level and high-level processes access and exchange information from bottom to top and from top to bottom.

It is to be remarked that the architecture is not to be considered as an empirical model of human perception. No hypotheses concerning its cognitive adequacy from a psychological point of view have been made; however, various cognitive results have been taken as sources of inspiration.

Preliminary results from an initial implementation of the described cognitive architecture are reported. The implementation has been developed by employing the

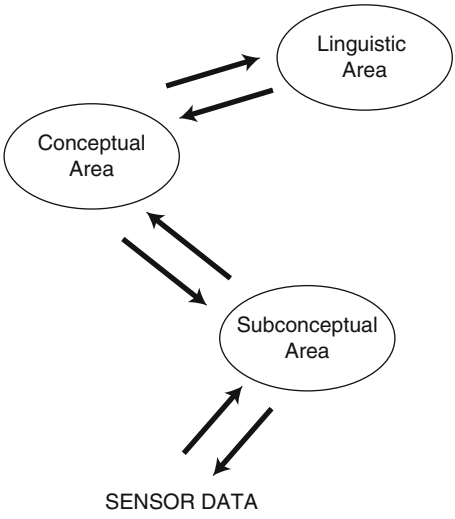

Fig. 10.1 A sketch of the cognitive architecture of the musical agent

Clozure Common Lisp,[1] the dissonance function developed by William Sethares[2] and the PowerLoom Knowledge Representation System[3] developed by the University of Southern California. In order to explore some characteristics of the adopted conceptual space, some common tools available in MATLAB[4] have been employed.

10.3 Conceptual Space of Musical Tones

A tone is a sound which is normally perceived in terms of pitch and timbre (Roederer 2008; Oxenham 2013). The pitch is related with the *highness* and *lowness* of the tone: for example the A is perceived as higher than C when they refer to the same musical scale. The timbre is related with the physical characteristics of the musical instrument: for example, the same C tone played by a piano and by a violin are perceived as different tones with the same highness but different timbre.

A *pure* tone is generated by a sine wave oscillator at a suitable frequency. For example, the pure tone A_4 (the A approximately at the middle of the piano keyboard) is generated by a sine wave at $f = 440$ Hz. Figure 10.2 shows the sound pressure (in arbitrary units) generated by a sinusoidal waveform at $f = 440$ Hz with amplitude 0.8. Figure 10.3 shows the resulting frequency spectrum.

According to the Discrete Fourier Analysis (DFT), a *complex* tone, as the one generated by a piano or a violin, may be decomposed as a superimposition of different sine waves at frequencies $f_1 = f; f_2 = 2f; f_3 = 3f; f_4 = 4f; \ldots$. The frequency f_1 is the *fundamental* frequency, the other frequencies f_i are multiples of the fundamental and they are called *partials*. Each i-th partial has an associated amplitude v_i. Figure 10.4 shows the sound pressure of a complex tone (in arbitrary units) generated by a fundamental sine wave at $f_1 = 440$ Hz and by superimposing sinusoidal waveforms at $2f, 3f, 4f$ with different amplitudes. Figure 10.5 shows the resulting frequency spectrum obtained by the Discrete Fourier Transform.

The aim of this section is to propose a conceptual space where a generic point $\mathbf{k}$ corresponds to a perceived tone. Therefore, a point $\mathbf{k}$ in the conceptual space will be expressed as the set of couples $((f_1, v_1)(f_2, v_2)\ldots)$ where f_i if the partial i of the tone ($i = 1$ is the fundamental), and v_i is the corresponding amplitude.

Considering the complex tone shown in Fig. 10.4, the corresponding point in the conceptual space will be represented as:

$$\mathbf{k} = ((440, 0.8)(880, 0.5)(1{,}320, 0.3)(1{,}760, 0.6)). \tag{10.1}$$

[1] http://ccl.clozure.com

[2] http://sethares.engr.wisc.edu/comprog.html

[3] http://www.isi.edu/isd/LOOM/PowerLoom

[4] http://www.mathworks.com

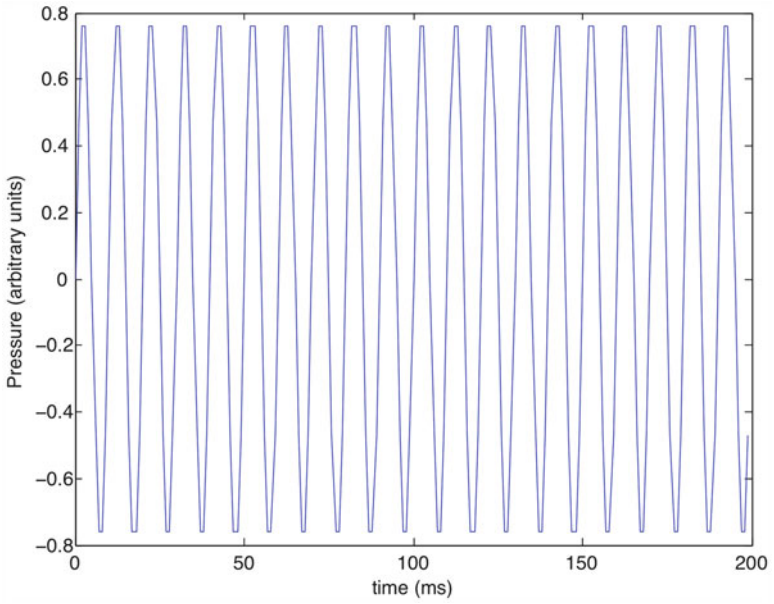

Fig. 10.2 The waveform of a pure tone

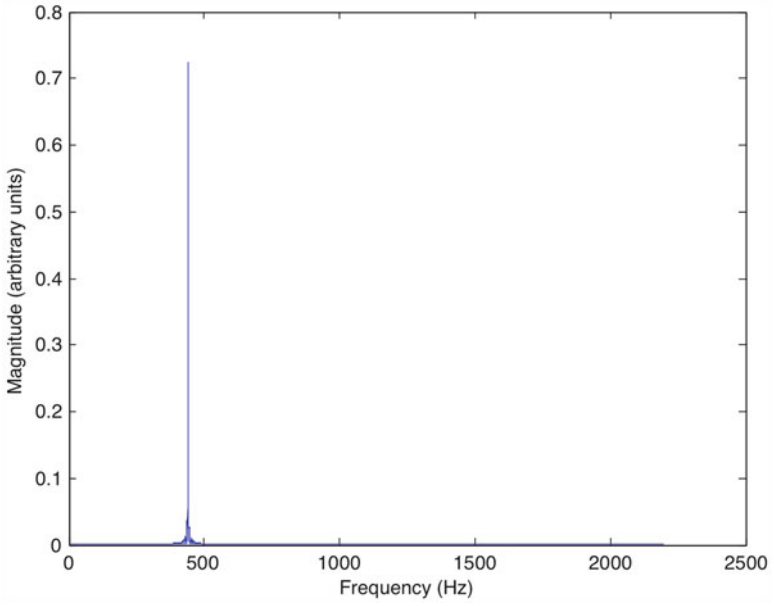

Fig. 10.3 The resulting spectrum of frequencies of the pure tone shown in Fig. 10.2

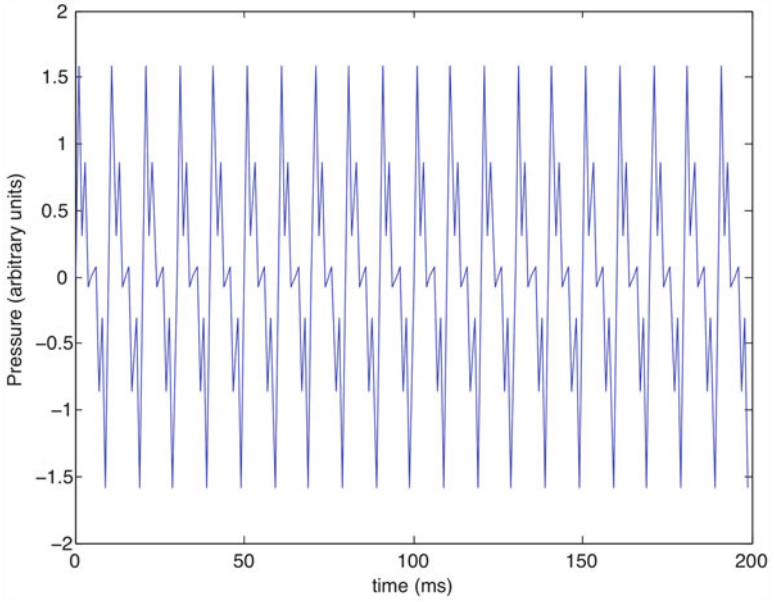

Fig. 10.4 The waveform of a complex tone

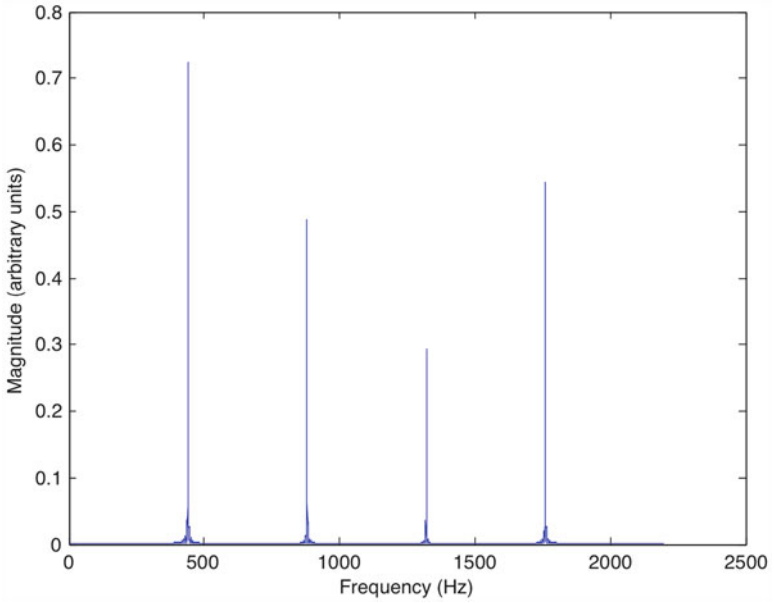

Fig. 10.5 The resulting spectrum of the complex tone shown in Fig. 10.4

From a mathematical point of view, the conceptual space so defined is a functional space whose basis functions are the trigonometric functions corresponding to the frequencies of the partials of the perceived tone. The coordinates of a point in this space thus correspond to the amplitudes of the partials resulting from the DFT of the tone.[5]

The quality dimensions of the conceptual space correspond to the partials of the perceived tone. They are *integral* dimensions in the sense of Gärdenfors (2000): in facts, the partials of a tone are typically not perceived as isolated sounds, but the whole tone is perceived as a unity.

In order to define the musical conceptual space, a distance measure d between tones is to be introduced. The distance measure, as discussed by Gärdenfors (2000), should capture a model of similarity, in the sense that two points in the conceptual space, corresponding to two tones perceived as similar, should have a smaller distance than two points corresponding to tones perceived as different.

In musical terms, two tones played simultaneously or in sequence is an *interval*. It is well known from Pythagoras that pleasant intervals played on a vibrating string are related to simple string length ratios, as the octave (2/1), the fifth (3/2), the fourth (4/3). These are typically perceived as consonant intervals and in facts they are the most common ones in Western music (Roederer 2008; Thompson 2013). However, the tendency for consonant intervals seems to be independent from culture, as it has been noticed in months-aged infants (Trehub 2001).

In the context of music perception, we propose to consider the similarity of two tones in terms of their *consonance*: i.e., two tones which are close in our conceptual space are perceived as consonant ones while two tones which are distant are perceived as dissonant ones. In other words, the similarity d between points in our conceptual space corresponds to the consonance of the musical interval of the two tones.

According to Terhardt (1974, 1984), consonance is a complex process generated by the intermixing of the *sensory* consonance related with psychoacoustics processes, and the *harmony* consonance related with the matching degree of the perceived tones and intervals with the *harmonic templates* acquired during life. Therefore, consonance is a mix of sensorial aspects and cultural aspects. In the following, we will mainly take into account the *sensory* aspects of consonance in the conceptual area, while we will consider cultural aspects of consonance in the linguistic area of the architecture.

As noticed by Helmholtz (1954), the consonance of an interval of two complex tones depends on the number of partials they have in common and on the distance between the other partials with respect to a suitable critical band. In facts, the octave and the fifth interval are the most consonant ones because they have many partials in common and the other partials do not interact.

[5]It should be noticed that this functional space has potentially infinite dimensions, but in practice the frequencies corresponding to the partials may be limited to a suitable range.

It is well known, since the seminal studies by Helmholtz, that consonance is associated with the absence of beatings. Beatings arise when the two tones generate constructive and destructive interferences. Qualitatively, when the difference Δf between the frequencies of two pure tones is near zero, then a smooth single tone with slow beatings is perceived. When the difference between the two frequencies grows up, then the perception of beatings disappears and the interval is perceived as dissonant and associated with a sense of roughness. As the difference between the two tones continues to grow up, the sensation of roughness disappears and the two tones are perceived as two separate tones. The transition from smoothness to roughness and dissonance and then to smoothness again depends on a *critical band* of frequencies (Zwicker et al. 1957) that in turn depends on the absolute values of the tones.

Plomp and Levelt (1965) performed an empirical analysis of consonance by means of subjective judgements of intervals made up by two pure tones; the resulting curve is characterized by a maximum peak at the starting frequency related with the unison interval, then a valley approximately before the second interval, then the curve rises again reaching a maximum peak approximately between the intervals of third and fourth, and eventually the curve slowly decreases. The width of the valley is related with the previously mentioned critical band. Extending the analysis to intervals of complex tones, they found that the intervals judged as consonant are, again, the intervals commonly employed in Western music: the unison, the octave, the fifth, the fourth (in order of consonance).

Several computational models have been proposed to model the consonance of pure and complex tones (Kameoka and Kuriyagawa 1969a,b; Hutchinson and Knopoff 1978; Mashinter 2006). In our architecture, we adopted the model proposed by Sethares (1993, 2005).

In details, Sethares parametrized the Plomp and Levelt curves by considering the following model:

$$d(f_1, f_2) = e^{-a(f_2 - f_1)} - e^{-b(f_2 - f_1)}. \tag{10.2}$$

where f_1 is the frequency of the first tone, f_2 is the frequency of the second tone and a and b are suitable model parameters. Figure 10.6 shows the dissonance curve generated by the algorithm when the starting frequency is $f = 440\,\text{Hz}$; in the Figure, the value 1 of the x axis corresponds to the unison interval, the value 1.5 corresponds approximatively to the fifth interval and the value 2 corresponds to the octave interval.

In order to model the dissonance between two complex tones related with two points $\mathbf{k_1}$ and $\mathbf{k_2}$ of the conceptual space, following Sethares we take into account an additive model that computes the dissonance of complex tones as the sum of the dissonances of the corresponding partials:

$$D(\mathbf{k_1}, \mathbf{k_2}) = \sum_{i=1}^{n} \sum_{j=1}^{n} d(f_i, v_i, f_j, v_j) \tag{10.3}$$

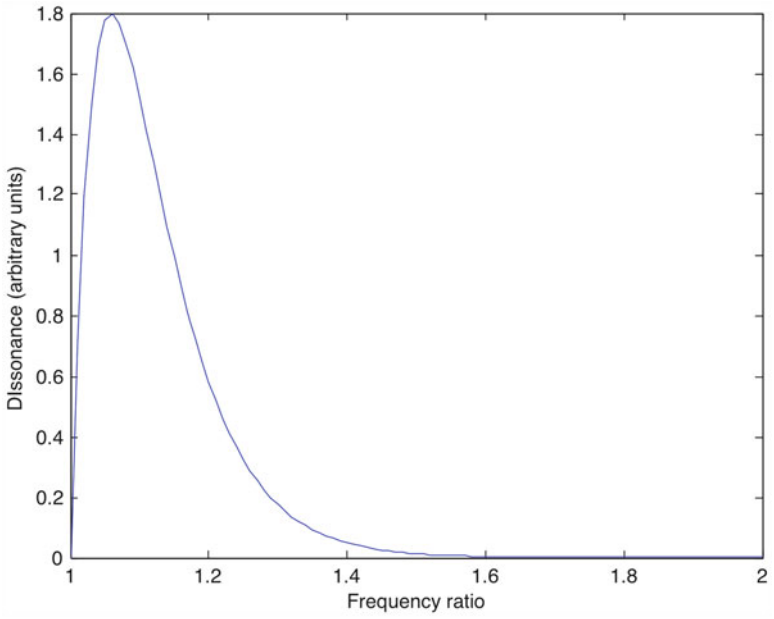

Fig. 10.6 The dissonance curve obtained by the Sethares model for pure tones. The value 1 of the x axis is the unison interval, the value 1.5 is approximatively the fifth interval and the value 2 is the octave interval

where (f_i, v_i) are the frequencies and amplitudes of the partials of $\mathbf{k_1}$, (f_j, v_j) are the frequencies and amplitudes of the partials of $\mathbf{k_2}$ and n is the sum of partials of the two tones; see Sethares (1993, 2005) for mathematical details.

As an example, Fig. 10.7 shows the dissonance curve obtained by the Sethares model by considering all the intervals generated by sweeping the complex tone depicted in Fig. 10.5 along the frequency range.

The minima of the dissonance curve correspond to the most consonant intervals. It is to be noticed from the figure that the adopted complex tone generates approximatively the following consonant intervals: unison (1/1), octave (2/1), fifth (3/2), fourth (4/3), major sixth (5/3), minor sixth (8/5), major third (5/4), minor third (6/5) (in order of consonance). Different choices for the amplitudes of the partials of the complex tone employed as seed may generate different orderings of intervals.

As another example, let us consider the following tones:

$$\mathbf{C_4} = ((261.6, 0.8)(523.2, 0.5)(784.8, 0.3)(1{,}046.4, 0.6)) \tag{10.4}$$

$$\mathbf{D\flat_4} = ((293.7, 0.8)(587.3, 0.5)(881, 0.3)(1{,}174.7, 0.6)) \tag{10.5}$$

$$\mathbf{G_4} = ((392, 0.8)(784, 0.5)(1{,}176, 0.3)(1{,}568, 0.6)). \tag{10.6}$$

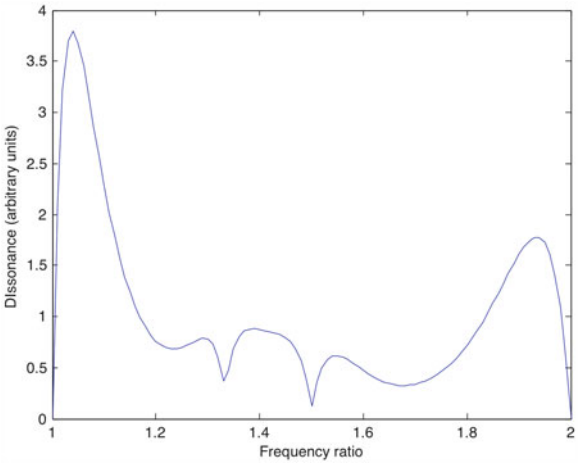

Fig. 10.7 The dissonance curve obtained by the Sethares model for the complex tone in Fig. 10.5. The value 1 of the x axis is the unison interval, the value 1.5 is approximatively the fifth interval and the value 2 is the octave interval

The dissonance measure between $\mathbf{C_4}$ and $\mathbf{D\flat_4}$ is related with a dissonant interval:

$$D(\mathbf{C_4}, \mathbf{D\flat_4}) = 2.56 \tag{10.7}$$

while the dissonance between $\mathbf{C_4}$ and $\mathbf{G_4}$ corresponds to the fifth interval which is second most consonant interval after the octave:

$$D(\mathbf{C_4}, \mathbf{G_4}) = 0.39 \tag{10.8}$$

Figure 10.8 is a pictorial representation of the discussed musical conceptual space, obtained by performing the Multidimensional Scaling Analysis on the dissonances related with the C4 major scale. As previously remarked, the most consonant intervals are the ones related with the fifth interval: $\mathbf{C_4} - \mathbf{G_4}$, $\mathbf{D_4} - \mathbf{A_4}$, $\mathbf{E_4} - \mathbf{B_4}$; the most dissonant ones are instead those one related with the second interval, the $\mathbf{C_4} - \mathbf{D_4}$, $\mathbf{D_4} - \mathbf{E_4}$, $\mathbf{E_4} - \mathbf{F_4}$, and so on.

The adopted dissonance measure may be also employed in order to analyze novel tones by means of their consonance with a set of prototypical tones. For example, let us consider a tone $\mathbf{k_1}$:

$$\mathbf{k_1} = ((523.2, 0.8)(784.8, 0.6)(1{,}046.4, 0.3)). \tag{10.9}$$

This tone is similar to $\mathbf{C_4}$ (Eq. 10.4) except for the fact that the fundamental frequency $f_1 = 261.6\,\text{Hz}$ is missing. The dissonance between $\mathbf{k_1}$ and $\mathbf{C_4}$ is:

$$D(\mathbf{C_4}, \mathbf{k_1}) = 0.04 \tag{10.10}$$

that allows the system to classify this tone as consonant with $\mathbf{C_4}$. As a comparison, the second best match is, as expected, $\mathbf{G_4}$, which corresponds to the fifth interval:

$$D(\mathbf{G_4}, \mathbf{k_1}) = 0.37. \tag{10.11}$$

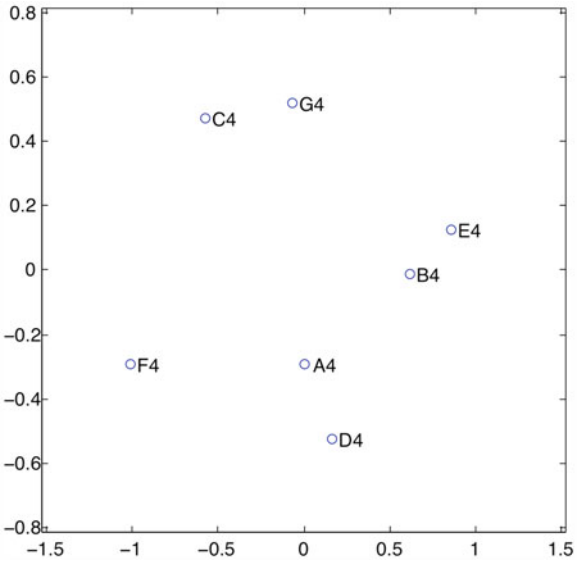

Fig. 10.8 A representation of the adopted conceptual space by means of Multidimensional Scaling

This is in line with the *virtual pitch* effect introduced by Terhardt (1974, 1984): a listener is able to perceive the pitch of a musical tone also when the fundamental frequency is missing. Terhardt hypothesizes a central pitch processor in the listener's brain that processes tones by a template matching process: during life, a listener is acquainted with many common music tones that allows her to learn several *harmonic templates*. When listening to a tone, the central pitch processor activates the acquired harmonic template that best fits the incoming tone also when the fundamental or other partials are missing (see also Roederer 2008).

As another example, let us consider the tone $\mathbf{k_2}$ obtained by $\mathbf{C_4}$ but characterized by many partials:

$$\begin{aligned}\mathbf{k_2} = &((261.6, 0.8)(523.2, 0.8)(784.8, 0.8)(1{,}046.4, 0.8)(1{,}308, 0.1)\\ &(1{,}569.6, 0.1)(1{,}831.2, 0.8)(2{,}092.8, 0.9)).\end{aligned} \tag{10.12}$$

Similarly, this tone is classified as consonant with $\mathbf{C_4}$, as it is in facts an instance of $\mathbf{C_4}$:

$$D(\mathbf{C_4}, \mathbf{k_2}) = 0.23 \tag{10.13}$$

The interval $\mathbf{A_4} - \mathbf{k_2}$ should therefore be a consonant interval, because it is a major sixth interval. Instead, in this case, the interval is a dissonant one:

$$D(\mathbf{A_4}, \mathbf{k_2}) = 2.45 \tag{10.14}$$

This is mainly due to the fact that some partials related with high frequencies of the tone $\mathbf{k_2}$ are close but not coinciding with the frequencies related with the partials of $\mathbf{A_4}$ thus generating roughness: in particular the frequency $f_5 = 1{,}308$ Hz in $\mathbf{k_2}$ is close to the frequency $f_3 = 1{,}320$ Hz in $\mathbf{A_4}$.

Again, see Sethares (1993, 2005) for detailed discussions on the adopted dissonance measure.

10.4 Linguistic Area

In the linguistic area of the architecture, the representation of perceived tones is based on a high level, logic oriented formalism. The linguistic area acts as a sort of long term memory, in the sense that it is a semantic network of symbols and their relationships related with musical perceptions. The linguistic area performs inferences of symbolic nature.

In the current implementation, the linguistic area is based on a knowledge representation system, namely the PowerLoom system (Chalupsky et al. 2010). A similar formalism in music perception has been adopted by Camurri et al. (1994, 1995).

The linguistic knowledge base encompasses a description of the tones and intervals that the architecture could encounter in its operations. The idea behind the mapping between the symbolic level and the conceptual level can be summarized as follows.

Single points in the conceptual space, i.e., pure and complex tones, correspond to individual constants at the linguistic level. As far as predicates are concerned, sets of points, e.g., sets of tones, correspond to one place predicates, while sets of pairs of points, e.g., sets of intervals, correspond to two place predicates.

This is not substantially different from what happens in usual model theoretic semantics for logic languages; what is novel with respect to traditional semantic techniques is that the conceptual space, compared with a traditional set theoretic interpretation, is endowed with a far more expressive structure. In facts, the conceptual level interpretation of the linguistic representations takes advantage from the geometric structure of the conceptual space itself.

In general, the description of the concepts in the symbolic knowledge base is not exhaustive. Only the information that is necessary in order to make suitable inferences is represented. By means of the interpretation of the symbolic knowledge base in terms of conceptual space structures, the linguistic area assigns names (symbols) to perceived entities and it describe their structure with a logical structural language. As a result, all the symbols in the linguistic area find their meaning in the conceptual space which is inside the system itself. A deeper account of these aspects can be found in Chella et al. (1997).

Figure 10.9 shows a fragment of the adopted knowledge base: a `tone` is an object characterized by at least one frequency and the relative amplitude, while a `complex-tone` is a tone characterized by several frequencies and related

```
(defconcept tone (?p thing))
(defrelation freq1 ((?p tone) (?f1 float)))
(defrelation amp1 ((?p tone) (?m1 float)))

(defconcept complex-tone (?p tone))
(defrelation freq2 ((?p complex-tone) (?f2 float)))
(defrelation amp2 ((?p complex-tone) (?m2 float)))
(defrelation freq3 ((?p complex-tone) (?f1 float)))
(defrelation amp3 ((?p complex-tone) (?m3 float)))
...
(defconcept C4 (?p complex-tone))
(assert (= (freq1 C4) 262))
(assert (= (freq2 C4) 523))
...
(assert (complex-tone k1))
(assert (complex-tone k2))

...
```

Fig. 10.9 A fragment of the knowledge base related with the definition of tones

```
(defconcept music-interval (?p thing))
(defrelation int1 ((?p music-interval) (?t tone)))
(defrelation int2 ((?p music-interval) (?t tone)))
(defrelation interval-type ((?p  music-interval)
                  (?t interval-class)))
(defrelation consonance ((?p music-interval)
                  (?x consonance-class)))
...
(assert (music-interval h1))
(assert (int1 h1 A4))
(assert (int2 h1 k2))
...
```

Fig. 10.10 A fragment of the knowledge base related with the definition of intervals

amplitudes. The tone `C4` is then a `complex-tone`. As stated in the previous section, `C4` is related with a prototypical point in the conceptual space. Then, `k1` and `k2` correspond to points in the conceptual space with coordinates defined respectively by Eqs. 10.9 and 10.12.

Figure 10.10 shows another fragment of the knowledge base related with the definition of intervals. A `music-interval` is an object characterized by two instances of the concept `tone`, an `interval-type` describing the type of the interval, e.g., if it is an unison, a fifth, a fourth, and so on. The elements of the `consonance-class` of interval are: `consonant` in the case of unison, octave, fifth and fourth intervals, `imperfect` in the case of sixth and third

```
(forall (?p music-interval)
        (=> (or (interval-type ?p M3)
                (interval-type ?p M6))
            (consonance ?p imperfect)))
```

Fig. 10.11 An example of a rule that classifies the major third and major sixth intervals as imperfect ones

intervals, and `dissonant` in all of the other cases. Then `h1` is an instance of `music-interval`, whose components are `A4` and `k2`.

As previously stated, the knowledge base is a repository that stores the default symbolic knowledge related with tones and intervals. In this sense, the linguistic area may also represent the cultural components of consonance as discussed by Terhardt (1974, 1984).

For example, in the case of intervals, a suitable symbolic rule as the one reported in Fig. 10.11, will classify by default a major sixth interval as an imperfect one. In particular, the interval `h1` (Fig. 10.10) will be classified as `imperfect` according to this rule. However, as discussed in the previous Section, the dissonance between $\mathbf{A_4}$ and $\mathbf{k_2}$ is quite high, and therefore it is a `dissonant` interval, after all. In this case, the dissonance measure allows the system to retract the previous symbolic default knowledge.

This is an example showing how the link between the conceptual space and the symbolic area may manage some non trivial mechanisms related with non monotonic logic in a simple way (Gärdenfors and Williams 2001).

10.5 Discussion and Conclusions

A cognitive architecture based on conceptual spaces has been outlined for an agent able to perceive music. In analogy with static perception of scenes, a conceptual space has been proposed aimed at representing tones and intervals. A measure is defined in the conceptual space that relates the similarity of tones with their consonance: roughly, two tones are similar if they are parts of a consonant interval.

Several problems arise concerning the proposed approach. First, the adopted dissonance measure is not a metric in the strict geometric sense; therefore more theoretical investigations are needed, in the line of e.g. the approach proposed by Tversky (1977).

A main practical problem, analogously with the case of computer vision, concerns *segmentation*. In the case of our conceptual space, the musical agent would be able to segment the partials of the signal coming from the microphone in order to individuate the frequencies of tones and intervals. Although many algorithms for music segmentation have been proposed in the computer music literature and some of them are also currently available in commercial programs, as for example the *AudioSculpt* program developed by IRCAM,[6] this is a main problem in perception.

[6] http://forumnet.ircam.fr/product/audiosculpt/

However, the proposed architecture can be employed to help face the segmentation problem: the linguistic information can provide interpretation contexts and hypotheses that may help in segmenting the audio signal. An example of this interplay is at the basis of the IPUS system for the understanding of sounds (Lesser et al. 1995).

Another problem is related with the analysis of *time*. Currently, the proposed architecture does not take into account the metrical structure of the perceived tones; this is left for future works. Interesting starting points that could well fit with the proposed architecture have been proposed by Forth et al. (2010) and by Bååth et al. (2013).

In the discussed musical conceptual space, the *timbre* is in some sense implicitly represented by means of the amplitudes of partials of the perceived tones. This is a too simplified approach to the complex notion of timbre. Also in this case, this problem is left for future works. However, a specific conceptual space for the representation of timbres could be adopted by taking into account the empirical results reviewed by McAdams (2013).

Actually, there is not a clear boundary between musical knowledge that has to be provided at the linguistic level and the knowledge that has to be learned by the system, by means, of e.g. neural networks related with the subconceptual area. However, some main directions are emerging in the debate concerning the cognitive principles of music, as the theory of tonal music by Lerdahl and Jackendoff (1983). On the other side, the previously cited MUSACT system provides useful insights of the structures of the neural networks able to learn and generate expectations about tones and intervals.

In summary, an intermediate level based on conceptual spaces could be a great help towards the integration of the music cognitive systems based on subsymbolic representations, and the systems based on symbolic models of knowledge representation and reasoning. In facts, conceptual spaces could offer a theoretically well founded approach to the grounding of symbolic musical knowledge.

Finally, the relationships between music and vision are multiple and multifaceted. Conceptual space could offer a principled framework to explore these synergies towards a novel, unified theory of perception.

References

Bååth, R., Lagerstedt, E., & Gärdenfors, P. (2013). An oscillator model of categorical rhythm perception. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society*, Berlin (pp. 1803–1808). Austin, TX: Cognitive Science Society.

Bharucha, J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception: An Interdisciplinary Journal, 5*(1), 1–30.

Bharucha, J. (1991). Pitch, harmony and neural nets: A psychological perspective. In P. Todd & D. Loy (Eds.), *Music and connectionism* (pp. 84–99). Cambridge: MIT.

Boden, M. (2004). *The creative mind: Myths and mechanisms* (2nd ed.). London: Routledge.

Bregman, A. (1994). *Auditory scene analysis – The perceptual organization of sound*. Cambridge: MIT/Bradford Book.

Camurri, A., Frixione, M., & Innocenti, C. (1994). A cognitive model and a knowledge representation system for music and multimedia. *Journal of New Music Research, 23*, 317–347.

Camurri, A., Catorcini, A., Innocenti, C., & Massari, A. (1995). Music and multimedia knowledge representation and reasoning: The HARP system. *Computer Music Journal, 19*(2), 34–58.

Chalupsky, H., MacGregor, R., & Russ, T. (2010). *PowerLoom manual*. Marina Del Rey: University of Southern California.

Chella, A., Frixione, M., & Gaglio, S. (1997). A cognitive architecture for artificial vision. *Artificial Intelligence, 89*, 73–111.

Chella, A., Frixione, M., & Gaglio, S. (1998). An architecture for autonomous agents exploiting conceptual representations. *Robotics and Autonomous Systems, 25*(3–4), 231–240.

Chella, A., Frixione, M., & Gaglio, S. (2000). Understanding dynamic scenes. *Artificial Intelligence, 123*, 89–132.

Chella, A., Gaglio, S., & Pirrone, R. (2001). Conceptual representations of actions for autonomous robots. *Robotics and Autonomous Systems, 34*, 251–263.

Chella, A., Frixione, M., & Gaglio, S. (2003). Anchoring symbols to conceptual spaces: the case of dynamic scenarios. *Robotics and Autonomous Systems, 43*(2–3), 175–188.

Chella, A., Frixione, M., & Gaglio, S. (2008). A cognitive architecture for robot self-consciousness. *Artificial Intelligence in Medicine, 44*, 147–154.

Clarke, E. (2005). *Ways of listening. An ecological approach to the perception of musical meaning*. Oxford: Oxford University Press.

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception, 32*, 341–365.

Deutsch, D. (2013). Grouping mechanisms in music. In: D. Deutsch (Ed.), *The psychology of music* (3rd ed., chap. 6, pp. 183–248). Amsterdam: Academic.

Forth, J., Wiggins, G., & McLean, A. (2010). Unifying conceptual spaces: Concept formation in musical creative systems. *Minds and Machines, 20*, 503–532.

Gärdenfors, P. (1988). Semantics, conceptual spaces and the dimensions of music. In: V. Rantala, L. Rowell, & E. Tarasti (Eds.), *Essays on the philosophy of music* (pp. 9–27) Helsinki: Philosophical Society of Finland.

Gärdenfors, P. (2000). *Conceptual spaces*. Bradford Books: MIT.

Gärdenfors, P., & Williams M. A. (2001). Reasoning about categories in conceptual spaces. In *Proceedings of the seventeenth international joint conference on artificial intelligence*, Seattle (pp. 385–392).

Gibson, J. (1979). *The ecological approach to visual perception.* Hillsdale: Lawwrence Erlbaum Associates.

Helmholtz, H. (1954). *On the sensations on tone as a physiological basis for the theory of music* (Alexander J. Ellis, Trans.). New York: Dover Publications, Inc.

Hutchinson, W., & Knopoff, L. (1978). The acoustic component of western consonance. *Interface, 7*(1), 1–29.

Kameoka, A., & Kuriyagawa, M. (1969a). Consonance theory, part I: Consonance of dyads. *Journal of the Acoustical Society of America, 45*(6), 1451–1459.

Kameoka, A., & Kuriyagawa, M. (1969b). Consonance theory, part II: Consonance of complex tones and its computation method. *Journal of the Acoustical Society of America, 45*(6), 1460–1469.

Krueger, J. (2009). Enacting musical experience. *Journal of Consciousness Studies, 16*, 98–123.

Leman, M. (2008). *Embodied music cognition and mediation technology.* Cambridge: MIT.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge: MIT.

Lesser, V., Nawab, H., & Klassner, F. (1995). IPUS: An architecture for the integrated processing and understanding of signals. *Artificial Intelligence, 77*, 129–171.

London, J. (2004). *Hearing in time: Psychological aspects of musical meter.* Oxford: Oxford University Press.

Marr, D. (1982). *Vision.* New York: W.H. Freeman.

Mashinter, K. (2006). Calculating sensory dissonance: Some discrepancies arising from the models of Kameoka & Kuriyagawa, and Hutchinson & Knopoff. *Empirical Musicology Review, 1*(2), 65–84.

McAdams, S. (2013). Musical timbre perception. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., chap. 2, pp. 35–67). Amsterdam: Academic.

Oxenham, A. (2013). The perception of musical tones. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., chap. 1, pp. 1–33). Amsterdam: Academic.

Pearce, M., & Wiggins, G. (2004). Improved methods for statistical modelling of monophonic music. *Journal of New Music Research, 33*(4), 367–385.

Pearce, M., & Wiggins, G. (2006). Expectation in melody: The influence of context and learning. *Music Perception: An Interdisciplinary Journal, 23*(5), 377–406.

Plomp, R., & Levelt, W. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America, 38*(4), 548–560.

Roederer, J. (2008). *The physics and psychophysics of music – An introduction* (4th ed.). Berlin: Springer.

Sethares, W. (1993). Local consonance and the relationship between timbre and scale. *Journal of the Acoustical Society of America, 94*(3), 1218–1228.

Sethares, W. (2005). *Tuning, timbre, spectrum, scale* (2nd ed.). London: Springer.

Tanguiane, A. (1993). *Artificial perception and music recognition* (Lecture notes in artificial intelligence, Vol. 746). Berlin/Heidelberg: Springer.

Temperley, D. (2001). *The cognition of basic musical structures.* Cambridge: MIT.

Temperley, D. (2012). Computational models of music cognition. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., chap. 8, pp. 327–368). Amsterdam: Academic.

Terhardt, E. (1974). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America, 55*(5), 1061–1069.

Terhardt, E. (1984). The concept of musical consonance: a link between music and psychoacoustics. *Music Perception, 1*(3), 276–295.

Thompson, W. (2013). Intervals and scales. In: D. Deutsch (Ed.), *The psychology of music* (3rd ed., chap. 4, pp. 107–140). Amsterdam: Academic.

Trehub, S. (2001). Musical predispositions in infancy. *Annals of the New York Academy of Sciences, 930*, 1–16.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327–352.

Wiggins, G. (2006). A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems, 19*, 449–458.

Wiggins, G., Pearce, M., & Müllensiefen, D. (2009). Computational modelling of music cognition and musical creativity. In R. Dean (Ed.), *The Oxford handbook of computer music* (chap. 19, pp. 387–414). Oxford: Oxford University Press.

Windsor, W., & de Bèzenac, C. (2012). Music and affordances. *Musicae Scientiae, 16*(1), 102–120.

Zwicker, E., Flottorp, G., & Stevens, S. (1957). Critical bandwidth in loudness summation. *Journal of the Acoustical Society of America, 29*, 548–557.

Part IV
Philosophical Perspectives

Chapter 11
Conceptual Spaces as Philosophers' Tools

Lieven Decock and Igor Douven

Abstract This paper gives an overview of the main philosophical applications to which conceptual spaces have been put. In particular, we show how they can be used to (i) resolve in a uniform way the so-called paradoxes of identity, which are basically problems concerning material constitution and change over time; (ii) answer one of the core questions in the debate concerning vagueness, to wit, the question of what a borderline case is, for instance, what makes some items neither clearly green nor clearly not green but borderline green; and, building on this answer, give a philosophically coherent account of the graded membership relation that is at the heart of fuzzy set theory; and (iii) provide a novel analysis of the concept of knowledge, which answers in a conservative way questions recently raised about the relationship between knowledge (or knowledge ascriptions) and the practical interests of putative knowers.

Over the past two decades or so, the conceptual spaces approach has been firmly established as one of the main accounts of conceptualization. The approach has its roots in empirical and theoretical work carried out by cognitive psychologists from the 1980s onwards, but it was only generally recognized as a new research paradigm after the publication of Peter Gärdenfors' *Conceptual Spaces* (Gärdenfors 2000).

The key tenet of the approach is that concepts – which, to begin, we may roughly characterize as the mental correlates of predicates – can be thought of *more geometrico*, as regions in mathematical spaces of a special kind. This is very different from the way philosophers have traditionally thought of concepts, to wit, as lists of conditions all or most of which must be satisfied by an item if that item is to fall under the concept. Nevertheless, philosophers can only be impressed by the empirical successes of the conceptual spaces approach and so, given the generally naturalistic outlook of many present-day philosophers, and given that the nature of

L. Decock (✉)
Department of Philosophy, VU University Amsterdam, Amsterdam, The Netherlands
e-mail: l.b.decock@vu.nl

I. Douven
Sciences, Normes, Décision, CNRS/Paris-Sorbonne University, Paris, France
e-mail: igor.douven@paris-sorbonne.fr

F. Zenker, P. Gärdenfors (eds.), *Applications of Conceptual Spaces*,
Synthese Library 359, DOI 10.1007/978-3-319-15021-5_11

concepts is one of the central questions in the philosophy of mind, it is fair to say that the approach deserves more attention from philosophers than it has received so far.

In fact, the approach offers more than just an interesting new account of concepts. A number of papers of our own, some written in collaboration with other researchers, illustrate some of the ways in which the conceptual spaces approach can be used to elucidate and sometimes even solve traditional philosophical problems from a variety of areas, including metaphysics, philosophy of logic, philosophy of language, and epistemology. For a long time, the main formal tool of analytic philosophers has been logic. More recently, probability theory has been added to the philosopher's toolbox, especially for those working in epistemology and the philosophy of science. The further addition of conceptual spaces to the toolbox makes some philosophical problems amenable to a *geometrical* treatment, as our previous research showed.

In the following, we give an overview of the main philosophical applications to which conceptual spaces have been put. In particular, we show how they can be used to (i) resolve in a uniform way the so-called paradoxes of identity, which are basically problems concerning material constitution and change over time (Sect. 11.1); (ii) answer one of the core questions in the debate concerning vagueness, to wit, the question of what a borderline case is, for instance, what makes some items neither clearly green nor clearly not green but borderline green; and, building on this answer, give a philosophically coherent account of the graded membership relation that is at the heart of fuzzy set theory (Sect. 11.2); and (iii) provide a novel analysis of the concept of knowledge, which answers in a conservative way questions recently raised about the relationship between knowledge (or knowledge ascriptions) and the practical interests of putative knowers (Sect. 11.3).

The purpose of this paper is not just to make more widely known the merits of the conceptual spaces approach as a tool for analytic philosophers. As will be seen, the applications of conceptual spaces to philosophical problems have in turn led to extensions of the conceptual spaces framework that should be of interest to cognitive scientists as well. Indeed, some of the philosophical work has not only added to the basic machinery available in the conceptual spaces framework, but also has clear empirical content, the testing of which should be of interest to philosophers and cognitive scientists alike.[1]

[1] In the following, we assume readers to be familiar with the basics of the conceptual spaces approach. Readers who are not may wish to consult the first chapters of Gärdenfors (2000) and the introduction to this volume.

11.1 Identity and Similarity

According to standard philosophical theorizing, the relation '_ is identical to _' is a perfectly unproblematic and simple relation: it is the smallest relation that is reflexive, symmetric, and transitive that obeys Leibniz's Law. That is to say, everything is identical to itself; if *a* is identical to *b*, then *b* is identical to *a*; if *a* is identical to *b* and *b* is identical to *c*, then *a* is identical to *c*; and if *a* and *b* are identical, then they have exactly the same properties.

However, many of our everyday ascriptions of identity seem at odds with at least some of these alleged properties of the identity predicate. Consider a famous puzzle case: We are inclined to judge as identical a statue and the piece of bronze that constitutes it, if only because it would be strange to hold that two different objects could occupy exactly the same space. On the other hand, we have no difficulty imagining events that would destroy the statue but not the piece of bronze. For instance, melting the bronze and creating a new statue out of it would destroy the original statue though, arguably, not the piece of bronze. But this means that the statue and the piece of bronze that we deemed identical do not share all their properties, and thus do not obey Leibniz's Law.

Here is an equally well-known case, one that was already discussed by the ancient Greek philosophers and that often goes by the name of 'paradox of Theseus' ship': We are inclined to judge that replacing just one plank from a ship does not turn the ship into a new one. Replacing a second plank is unlikely to alter that judgment. However, if we continue to replace planks, we will reach a point at which we no longer deem the resulting ship identical to the one with which we started. But then where does this leave the supposed transitivity of the identity predicate? If replacing a single plank does not deliver a new ship, then the ship with one plank replaced should be identical to the ship with zero planks replaced, the ship with two planks replaced should be identical to the ship with one plank replaced, and so on, which, if transitivity of the identity predicate is assumed, yields the conclusion that the ship with all planks replaced is identical to the ship with zero planks replaced.

Various other puzzle cases apparently involving the identity predicate are known in the literature that, jointly, seem to put considerable pressure on the simple view of identity cited above (henceforth, 'the simple view of identity'). In addition to seemingly showing that the identity relation does not (always) obey Leibniz's Law and is not transitive, these cases have been interpreted as giving reason to believe that the identity relation is context-sensitive, subjective, and (sometimes) vague. The puzzle cases are commonly known as 'the paradoxes of identity'.

Few philosophers have been willing to abandon the simple view of identity and have tried to diagnose the paradoxes of identity as signaling some confusion on our part about concepts other than identity, like the concept of an object or that of a property. It is fair to say, though, that the detailed proposals of how to deal with the various puzzle cases form a motley bunch in that they solve one puzzle in one way and another puzzle in quite a different way. For example, the paradox of the statue and the lump of bronze is typically dealt with in terms of constitution, while many

think that the paradox of Theseus' ship is to be solved by invoking the conceptual apparatus of mereology (i.e., the theory of parts and wholes).[2]

In Douven and Decock (2010), we set out to give a uniform solution to the paradoxes of identity. Just as is commonly believed to hold for scientific theories, we believe that for philosophical theories, the power to explain what at first seem to be disparate phenomena by reference to a single mechanism (or at least to no more than a few mechanisms) is a virtue which gives reason to believe that the theory is headed along the right lines. In the solution we proposed, the framework of conceptual spaces occupies center stage as a unifying mechanism.

The core idea of our 2010 paper is that the word 'identity' is ambiguous, positioned between the notion that figures in the simple view – basically, the notion of identity familiar from introductory logic courses – and a notion that is to be understood in terms of relevant similarity. According to our diagnosis in that paper, many of our everyday identity claims really involve the latter notion. What is more, we argue that this is also the notion that is involved in the paradoxes of identity – which on our account then only appear paradoxical because we mistake the relevant similarity notion of identity at play in them for the logical notion of the simple view.

More specifically, we argue in that paper that many identity claims are really claims to the effect that two things are highly similar in all relevant respects. The restriction to *relevant* respects is crucial here. There is a welter of psychological research showing that when we compare items with each other, we typically take into account only a subset of the respects in which the objects *could* be found to be similar, and that it is a context-dependent matter which subset we take into account. In light of this, the proposal that 'identity' is ambiguous and often means 'high similarity in all relevant respects' makes it unsurprising that our identity judgments can vary with context. In a context in which we attend to the different modal properties of a statue and the lump of bronze of which it is made, we may judge the statue and the lump of bronze to be non-identical; in another context, one in which those properties are not salient, we may deem the statue and the lump of bronze to be identical.

The proposal also makes it unmysterious how there can be cases of vague identity: whether two objects are highly similar in all relevant respects can easily be a vague matter, if only because of the vagueness that attaches, or at least can attach, to what counts as highly similar and which respects are relevant (that is to say, the vagueness that comes with the words 'high' and 'relevant').

In relation specifically to the paradox of Theseus' ship, it is to be noted that similarity is not generally transitive: a may be highly similar to b and b may be highly similar to c while a is not highly similar to c. Applied to the paradox: a ship with one plank replaced may be highly similar in all relevant respects – and in that sense identical – to the original ship; the ship with two planks replaced may be highly similar to the ship with only one plank replaced and may in that sense be identical to it; and so on for the rest of the series if we continue to replace planks.

[2]See Douven and Decock (2010) for references and further details.

But ships far enough apart from one another in this series may not be highly similar in all relevant respect, and thus may fail to qualify as identical.

Even if the notion of identity-as-high-similarity is not quite as simple as the logical or 'simple' notion of identity, in Douven and Decock (2010) we show that it can be defined in an equally precise fashion by drawing on the machinery of the conceptual spaces framework. The definition is presented in two steps.

In the first step, we equate a contextually relevant respect with a conceptual space that is being activated in the given context and further equate a context C with the set S_C of conceptual spaces activated in it. So, for instance, if color is a relevant respect of comparison in C, then S_C includes a color space; if shape is relevant in C, then S_C contains a shape space; etcetera.

As for the second step, let $d_r(\cdot,\cdot)$ be the metric defined on space $r \in S_C$, and let ${\mathbf{t}_C}^r \geqslant 0$ be a threshold associated with r. Then the formal definition of identity-as-high-similarity is this:

$$\text{(ID)} \quad \mathrm{Id}_C\,(a,b) \iff \forall r \in S_C : d_r\,(a,b) \leqslant {\mathbf{t}_C}^r.$$

If this is expressed in words, to be identical-as-highly-similar in a context C, a and b must in all conceptual spaces activated in C – all respects of comparison that are attended to, or relevant, in C – lie sufficiently close to each other,[3] where what counts as sufficiently close may, as the sub- and superscript of the threshold indicate, vary from one context to another as well as, within the same context, from one respect to another.

Shifts in context may make an aspect of comparison that was earlier irrelevant suddenly relevant or the other way round; or they may push upwards or downwards the thresholds associated with some of the activated spaces; or they may bring about simultaneously both of the foregoing types of changes. Another important fact to note is that, as Gärdenfors (2000: 89) shows, psychological measures, such as the metrics defined on conceptual spaces, are generally imprecise. And finally, that a and b lie 'close' to each other in a given space, and that the same holds for b and c, does not guarantee that a and c lie close to each other in that space. As we argue in our 2010 paper, in light of these three facts, (ID) allows us to account for the paradoxes of identity in a uniform way. Specifically, these facts account for the context-sensitivity of the identity-as-high-similarity relation, the vagueness that may attach to this relation, and its not being transitive, respectively.

For later purposes, we mention that (ID) can be easily extended so as to make it pertain to properties and not just to objects. The relevant point to notice is that it is possible to measure distances between *sets* of points in a conceptual space in much the same way in which distances between points are measured. The standard (even if not the only) metric for this purpose is the so-called Hausdorff metric, according to which the distance between sets Φ and Ψ equals

[3] To be exact, the points representing these objects in the various spaces must be sufficiently close to each other. We will sometimes leave this distinction implicit.

$$\max\left\{\sup_{x\in\Phi}\inf_{y\in\Psi}\delta_S(x,y),\sup_{y\in\Psi}\inf_{x\in\Phi}\delta_S(x,y)\right\}.$$

To put this more informally, one considers, for any point in Ψ, the shortest distances between that point and the points in Φ, and also considers, for any point in Φ, the shortest distances between that point and the points in Ψ, and then takes the longest of all those 'shortest distances', which is the Hausdorff distance between the two sets of points. This metric, and the generalization of (ID) that builds on it, will become relevant in Sect. 11.3, when we consider the application of the conceptual spaces approach to the concept of knowledge.

11.2 Vagueness, Borderline Cases, and Graded Membership

For almost any predicate in our language, we can think of cases such that the predicate does not quite apply to them but also does not quite apply *not* to them. Such cases are generally called 'borderline cases' (of the given predicate). We can imagine how chemists and physiologists may jointly succeed in giving a precise answer to the question of what makes the taste of a particular candy bar a borderline case of sweetness. But philosophers and formal semanticists have sought to characterize borderline cases generally: what is it that is shared by all borderline green/blue cases, borderline sweet/sour cases, borderline loud/not-loud cases, and so on?

Gärdenfors (2000) presents a version of the conceptual spaces approach that suggests a very simple and natural answer to the above question. The version equips each conceptual space with certain designated points and then applies the mathematical technique of Voronoi tessellations (or Voronoi diagrams) to obtain a segmentation of the space. The designated points represent the most typical instances of the concepts represented in the space; that is to say, they represent the concepts' prototypes, in the sense of Rosch (1975). For example, in color space – a three-dimensional mathematical structure, with the dimensions representing hue, luminosity, and saturation, approximately of the form of Fig. 11.1 – we find a point representing prototypical red, one representing prototypical blue, and so on for the other colors. Given such 'prototypical points' in a space, we can obtain a division of that space by associating a cell with each prototypical point and grouping all points in the space according to the principle that any point in any given cell must lie at least as close to the prototypical point with which the cell is associated as to any of the other prototypical points in the space. Figure 11.2 makes this a bit less abstract, by showing a two-dimensional space with designated points carved up by a Voronoi diagram generated by those points.

The idea is that concepts can be identified with cells in a space segmented by a Voronoi diagram generated by the prototypical points in that space. So if we imagine, for now, that color space is two-dimensional and represented by the space in Fig. 11.2, then the points in that space can be thought of as representing the

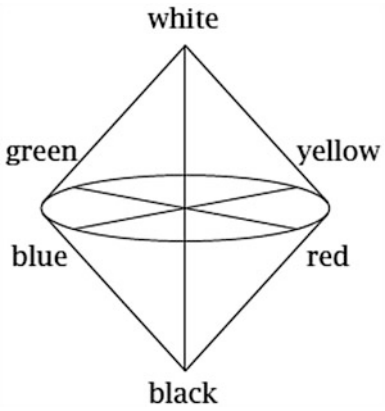

Fig. 11.1 Schematic representation of color space

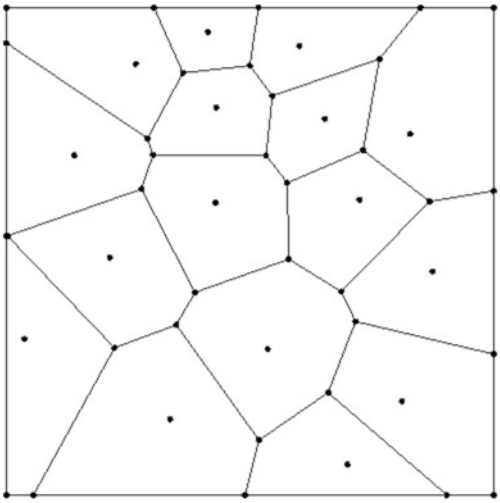

Fig. 11.2 Two-dimensional space with Voronoi diagram

most typical shades of the various colors, and the cells as representing the various color concepts. Figure 11.2 illustrates a nice feature of this way of thinking about conceptualization, to wit, that concepts are represented by *convex* regions in spaces, meaning that for any two items *a* and *b* falling within a concept, the items that lie in an intuitive sense 'between' *a* and *b* also fall within the same concept. For example, if we move along a straight line in color space from one shade of red to another shade of red, we only come across shades of red, and not of any other color. As Gärdenfors (2000) shows, there is considerable empirical support for this convexity assumption. Indeed, Gärdenfors is generally sympathetic to the conceptual-spaces-*cum*-prototypes-*cum*-Voronoi-diagrams framework, even if in the end he does not commit to it.

We now return to the question of how to give a general characterization of borderline cases. If we present this question to students while also introducing them

to the conceptual spaces framework and showing an illustration such as Fig. 11.2, and assuming the students not to be completely unimaginative, it is highly probable that the students will come up with the following answer: a borderline case of a concept *X* is a case that is represented by a point in the relevant conceptual space that is just as far from the point representing the *X*-prototype in the space as it is from a point representing the prototype of another concept (of which the case is then also a borderline case). Indeed, it seems that looking at Fig. 11.2, one cannot miss the borderlines between the concepts. And what is more natural than thinking that borderline cases are precisely the cases represented by points lying on such a borderline?

However natural and straightforward the answer may appear, it cannot be quite right. In particular, there are aspects of the phenomenology of borderline cases that it fails to capture. Imagine a color patch that is prototypically blue and then imagine that we change its color in a continuous fashion toward green, until its color is prototypically green. If color space were two-dimensional, this process could be imagined as going in a straight line from one designated point in Fig. 11.2 to a neighboring designated point in that figure. It is not difficult to see that in the process, we do not suddenly, after seeing a patch that is definitely blue, encounter a borderline blue/green patch, after which we immediately see a definitely green patch. Rather, we will be in a kind of borderline region for some while, seeing only borderline blue/green patches, even if borderline patches that occur later in the series will appear to be greener than borderline patches that occur earlier. So, from a phenomenological perspective, the borderlines that result from the conceptual spaces-*cum*-prototypes-*cum*-Voronoi-diagrams model are too 'thin'. In this picture, 'almost all' points neighboring a borderline point (i.e., a point representing a borderline case) are *not* themselves borderline points but belong definitely to one or the other concept, while in reality borderline points can be completely surrounded by borderline points. For instance, we can easily imagine a borderline case of blue/green such that slightly changing its color along any dimension in any direction will again result in a borderline blue/green case.

A solution to this problem was suggested by taking seriously an observation about the conceptual spaces-*cum*-prototypes-*cum*-Voronoi-diagrams model that at first appears entirely unrelated to the problem. The observation is that for many concepts, it can at best be an idealization to assume that there is a unique prototype of the concept. There are many shades of red that are clearly red but that are not typical for the color. For example, shades of red that we would call 'crimson' are clearly red and not any other color, but they are not typical for the color red. However, there are also many shades of red that, although slightly different from each other, all strike us as being typically red. This seems clear on the basis of a priori reflection, but it is also supported by empirical data: in their famous color naming experiment, Berlin and Kay (1969) found that their participants often designated more than one Munsell color chip if they were asked to point at the chip or chips they regarded as typical for a given color. In Douven et al. (2013), we concluded from this that, rather than with prototypical *points*, conceptual spaces

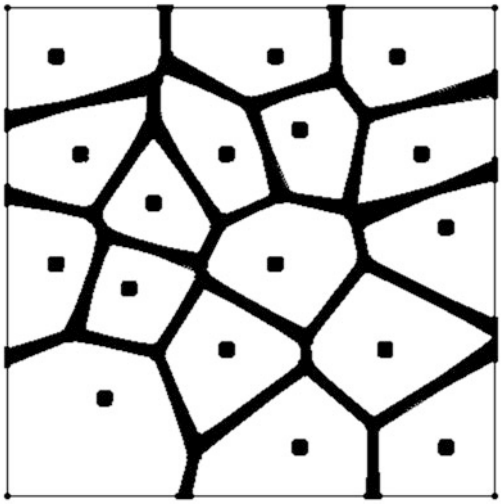

Fig. 11.3 Two-dimensional space with collated Voronoi diagram

are to be equipped with prototypical *regions*: regions that represent not the most representative *case* of a concept, but its most representative *cases*.

This raised the question of what remains of the other element of the model, the technique of Voronoi diagrams used to carve up the model on the basis of the representations of prototypes. Standard Voronoi diagrams are defined for sets of single generating points, not sets of regions. The modification of the technique proposed in Douven et al. (2013) is at bottom very simple. This modification starts by considering the set of all possible selections of one single point from each prototypical region in a space, and noting that each element of that set can be used to generate a Voronoi diagram of the space. Call the set of Voronoi diagrams of a space S that are thus generated V_S. The modification then constructs a new type of Voronoi diagram – called 'collated Voronoi diagram' – by projecting all the Voronoi diagrams in V_S onto each other, so to speak. It is shown in Douven et al. (2013) that the collated Voronoi diagram of a space still carves up the space in such a way that the regions representing the concepts are convex. The great advantage of the collated construction over simple Voronoi diagrams is that the former have 'thick' boundary regions. As a result, 'almost all' borderline cases find themselves surrounded by other borderline cases, as required. See Fig. 11.3 for an illustration.

This construction allows one to state precisely what the boundary region of a given concept is (see Douven et al. (2013) for details). The answer to our question of what a borderline case is then as follows: a borderline case of a given concept is a case that is represented by a point lying in the boundary region of the concept.

It might be thought that, while the new model accounts more adequately for the nature of borderline cases than does the model with simple Voronoi diagrams, it is phenomenologically speaking still not entirely adequate. As is readily seen in Fig. 11.3, in the new model we still have a seemingly abrupt transition from clear cases to borderline cases, in that the borderline region *sharply* delineates the

concepts. In reality, we tend to sense no such sharp dividing line between the clear cases and the borderline cases of a concept. Indeed, where the clear cases of a concept end and its borderline cases begin is, for most concepts, itself a vague matter. This is sometimes called 'the problem of higher-order vagueness'.

Here it is important to recall the earlier remark about the imprecision of psychological metrics (see Sect. 11.1). That by itself is a cause of vagueness: we cannot delineate sharply the clear cases of a concept from its borderline cases because we cannot sharply delineate *anything* in a conceptual space: it is as if we were measuring distances in such a space with a measuring rod whose ends can only be dimly perceived.

But it could be argued that there is more to higher-order vagueness than just the lack of a sharp dividing line between clear and borderline cases. It is not just that we experience no precise point at which we transition from the clear cases to the borderline cases; the transition is also *smooth*. When we travel in color space in a straight line from a clearly green case to a clearly blue one (i.e., we imagine seeing the color shades which lie on a straight line in color space connecting a clearly green and a clearly blue shade), we first encounter borderline green/blue cases that in an intuitive sense are still very close to the clearly green cases, and then, very gradually, the cases become closer to the clearly blue ones. This smoothness in our experience is not satisfactorily explained by reference to the imprecision of psychological metrics. In fact, we believe it is best explained in terms of the account of graded membership that we developed on the basis of the above account of borderline vagueness.

The account of graded membership did in fact have a different motivation, which was related to fuzzy logic and fuzzy set theory. Fuzzy logic and fuzzy set theory have been practically very successful branches of twentieth-century logic. Engineers have applied fuzzy logic in electronic devices ranging from washing machines to pocket calculators. And yet the mathematical logic community at large has remained skeptical about fuzzy logic, the standard complaint being that a coherent interpretation of the fuzzy membership relation is still lacking. Indeed, it is not straightforward to explicate what it might mean that, for instance, a certain object is only to some non-extreme degree a member of the set of red things.

In Decock and Douven (2014), our primary aim was to clarify the fuzzy membership relation by defining it in operational terms, that is, by proposing an empirically-grounded method for measuring to what degree a given item falls within a given concept. The measure for graded membership that we proposed relies on the extended conceptual spaces framework that was summarized above.

The measure is easiest to describe for the case – possibly non-existent in reality – in which each prototypical region in a space consists of only finitely many points. By way of concrete example, consider the two-dimensional space depicted in Fig. 11.4. That space contains four prototypical regions, each of which consists of two points $\{a, b\}$, $\{c, d\}$, $\{e, f\}$, and $\{g, h\}$. If we call this space again S, the set V_S has $2^4 = 16$ members. Given the proposal made in Decock and Douven (2014), a point within in this space belongs to a concept C to a degree that equals the number of members of V_S that locate the point in the cell associated with one of the two prototypical

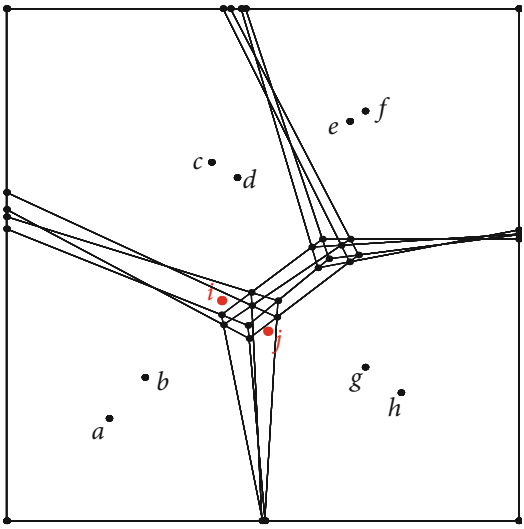

Fig. 11.4 Collated Voronoi diagram generated by four prototypical regions consisting of two points each, and designated points *i* and *j*

points *C* divided by the total number of members of V_S. For example, as is shown in our earlier paper, the point *i* belongs to the concept with the prototypical region {*a*, *b*} to a degree of ½, for 8 of the 16 simple Voronoi diagrams that make up the collated diagram of this space locate *i* in the cell associated with a member of {*a*, *b*}. Similarly, *j* belongs to the same concept to a degree of ¼.

The above generalizes swiftly to all conceptual spaces whose prototypical regions consist of a finite number of points. More importantly, it also generalizes to conceptual spaces whose prototypical regions consist of infinitely many points, but this generalization is not straightforward and requires some measure theory.[4]

For present purposes, the technical details of the measure–theoretic construction need not detain us. What is important is that, given that construction, it can be shown that membership functions are, in a clear sense, S-shaped: The degree of membership is 1 for the clear cases of a concept, 0 for the clear non-cases, and for borderline cases it gradually tapers off, in the form of an S, from 1 to 0 as we move away from the prototypical area. Not only is this in accordance with extant empirical results on membership functions of a number of concepts (e.g., Hampton (2007)); it also solves the problem concerning higher-order vagueness that we encountered earlier.

The problem was that, in human experience, there is a smooth transition from clear cases to borderline cases to clear non-cases, while in our model there are sharp categorical distinctions between, on the one hand, the clear cases and the borderline cases and, on the other, the borderline cases and the clear non-cases, as Fig. 11.3 suggests. Exactly where, in a space, the clear cases of a given concept in that space

[4]The resulting theory of graded membership generalizes further still to an account of graded truth, as shown in Douven and Decock (2015).

end and the borderline cases of the concept begin may be hard to tell due to the imprecision of the metric defined on the space; similarly, this is so for the boundary between the borderline cases and the clear non-cases. But that we have difficulty precisely locating these boundaries is not enough to explain the aforementioned experience of a smooth transition in going from the clear cases via the borderline cases to the clear non-cases.

This is where the results regarding the graded membership function are helpful. We experience the transition as smooth because it *is* smooth. That there is a sharp categorical distinction between the different types of cases – clear cases, borderline cases, and clear non-cases – means that there is an exact point (which we have difficulty *locating* exactly) at which the degree of membership of the cases we encounter is no longer 1. But at that point, the drop in degree of membership is so slight that we fail to experience it as engendering any kind of jolt; similarly, this is so when we leave the borderline region and encounter the first clear non-cases.

11.3 Knowledge: From Conceptual Analysis to Conceptual Construction

The conceptual spaces approach has been most extensively employed for modeling perceptual concepts, like color concepts, taste concepts, and olfactory concepts. It is by no means restricted to such concepts, though what exactly the scope of the approach is remains to be seen, and can only become clearer by trying to apply it to as great a diversity of concepts as possible. A recent success in this regard is Gärdenfors and Zenker's (2011), (2013) application of the conceptual spaces approach to fundamental scientific concepts – which for the most part are unarguably non-perceptual – in order to arrive at a natural account of scientific progress and theory change.

In Decock et al. (2014), the conceptual spaces approach is applied to another fundamental and obviously non-perceptual concept, which is central to epistemology, to wit, the concept of knowledge. The starting point of that paper is the by now common observation that our willingness to ascribe knowledge of a given proposition to a given person seems to depend, inter alia, on the *practical* value for that person of being right in believing the proposition. Thus, it seems, to answer the question of whether someone knows a proposition, we must go beyond asking strictly 'intellectualist' (i.e., truth related) questions, such as whether the proposition is true, whether the person believes the proposition to be true, or what is the quality of the person's evidence in favor of the truth of that proposition.

It does not follow, however, that the concept of knowledge involves any pragmatic elements. The truth or falsity of a knowledge claim might supervene on strictly truth-related factors. The apparent sensitivity of our knowledge ascriptions to what is practically at stake for the putative knower might then have a broadly Gricean explanation: in determining whether we are warranted in asserting that a

person knows a proposition, we might take into account the stakes for that person in being right about the proposition.

Yet an increasing number of authors think otherwise and hold that practical factors go into the truth conditions of knowledge-ascribing sentences. The purpose of Decock et al. (2014) is to show that the data about knowledge ascriptions warrant no such drastic departure from tradition (leaving it open that the departure may be warranted on other grounds) and can be explained by assuming that the concept of knowledge involves only intellectualist elements.

To show this, Decock et al. devise a conceptual space in which knowledge can be represented geometrically. Specifically, they take as dimensions *truth*, *belief*, and *justification*, though they also point out how further truth-related factors could be added as dimensions without altering in essence their explanation of the data concerning knowledge attributions that are at issue. In this space, knowledge occupies a fixed region and is thus insensitive to contextual variation, be it contextual variation related to changes in what is at stake or contextual variation due to other changes. Decock et al. then apply the theory of identity-as-high-similarity summarized in Sect. 11.1 to account for the sensitivity to variation in stakes of our willingness to attribute knowledge to people.

Their account crucially involves the notion of possessing at least approximate knowledge. Consider that the Eiffel Tower is 317 m high. Intuitively, someone who believes the Eiffel Tower to be 316 m high is closer to knowing the height of the Eiffel Tower than someone who believes the Eiffel Tower to be 315 m high, all else being equal. The same holds if someone's evidence for believing the Eiffel Tower to be 317 m is of better quality than another person's evidence for believing the same proposition, again ceteris paribus. It makes equally good intuitive sense to say non-comparatively that someone who believes the Eiffel Tower to be 316 m high and has excellent grounds for that belief is close to knowing the height of the Eiffel Tower; such a person has approximate knowledge of that height. And again the same holds if a person believes the Eiffel Tower is 317 m high but his or her evidence is just not good enough to yield knowledge.

Decock et al. show that, in the conceptual space in which knowledge is represented, the concept of at least approximate knowledge is a region that encompasses (the region representing) the concept of knowledge but occupies some extra space. Now recall from Sect. 11.1 that, on the identity-as-high-similarity view, identity is context-sensitive: whether two things qualify as being identical – for instance, a statue and the lump of bronze that constitutes it – may vary from one context to another. Further recall from that section that this view of identity applies equally to properties and concepts: whether two concepts qualify as being identical may depend on the context in which we ask this question. Decock et al. show that this applies to the concepts of knowledge and at least approximate knowledge. Specifically, they show that whether it is worth distinguishing between these two concepts depends on what is at stake for the putative knower. If not much is at stake, we can safely identify with each other knowledge and at least approximate knowledge: whether the putative knower knows a proposition, strictly speaking, or only knows that proposition approximately is not likely to make a difference for

any decision that he or she might base on that proposition. This is different in high-stakes contexts: in such contexts, it pays to finely distinguish between knowledge and approximate knowledge, for whether someone knows or only approximately knows is more likely to matter to what decision or decisions he or she might make.

The upshot is an explanation of the data at issue – the apparent sensitivity of knowledge attributions to practical concerns – that leaves intact the received view of knowledge as depending only on truth-related factors and that is non-ad hoc insofar as it follows from an application of an account of concepts in conjunction with a view of identity that were both developed for entirely unrelated reasons.

There is also a broader upshot, pertaining to the methodology of epistemology. Most, and probably even all, philosophers have been told in their introductory epistemology course that conceptual analysis is the single most important method applied in epistemology. Typically, such an analysis attempts to unravel a concept in terms of the conditions that are individually necessary and jointly sufficient for a thing to fall within the concept. The value of this method, when applied judiciously, is hard to overestimate, in epistemology as well as in other domains of analytic philosophy. But by regarding conceptual analysis as the only viable research tool, as is not uncommon among epistemologists, one may miss important avenues for making progress.[5]

Bechtel (2011) argues that biologists often try to explain biological phenomena in mechanistic terms. Mechanistic explanations crucially involve the identification and decomposition of the mechanism or mechanisms putatively giving rise to the phenomena. As Bechtel points out, however, the process of explanation cannot stop there; a full understanding of the phenomena requires more than a catalogue of the component parts of a mechanism. We must in addition know how these parts hang together, how they interact with one another, how they form a whole. To acquire such knowledge, biologists engage in what Bechtel calls 'recomposition', the process of constructing a model of the mechanism from the parts identified in the decomposition process.

If the conceptual spaces approach is moving along the right lines for at least many concepts, including the concept of knowledge, then the investigation of many concepts must involve more than an analysis of their component parts, and must involve some construction work as well: out of the component parts, a model of the concept must be built. That is the kind of work undertaken in Decock et al., and it was seen that it leads to a satisfactory explanation of the data about knowledge attributions. There is no reason to believe that the same approach will not work equally well for other philosophical concepts. Of course, only further research

[5] Some may want to go further still in claiming that the kind of conceptual construction typical of the conceptual spaces approach is superior to the traditional method of conceptual analysis. For instance, according to an anonymous referee, the conceptual spaces approach is "more cognitively sound than any conceptual analysis which relies mainly on two 'tools': intuition and logic. The greater cognitive adequacy of [conceptual spaces] for explaining concept formation and stabilization is compelling ... because it reconciles empirical data with philosophical analysis."

can tell. We hope that philosopher-colleagues will be inspired to undertake such research by means of the various applications of the conceptual spaces framework to philosophical issues that were highlighted above.[6]

References

Bechtel, W. (2011). Mechanism and biological explanation. *Philosophy of Science, 78*, 533–557.
Berlin, B., & Kay, P. (1969). *Basic color terms*. Berkeley: University of California Press.
Decock, L., & Douven, I. (2014). What is graded membership? *Noûs, 48*, 653–682.
Decock, L., Douven, I., Kelp, C., & Wenmackers, S. (2014). Knowledge and approximate knowledge. *Erkenntnis, 79*, 1129–1150.
Douven, I., & Decock, L. (2010). Identity and similarity. *Philosophical Studies, 151*, 59–78.
Douven, I., & Decock, L. (2015). What verities may be. *Mind*, in press.
Douven, I., Decock, L., Dietz, R., & Egré, P. (2013). Vagueness: A conceptual spaces approach. *Journal of Philosophical Logic, 42*, 137–160.
Gärdenfors, P. (2000). *Conceptual spaces*. Cambridge, MA: MIT Press.
Gärdenfors, P., & Zenker, F. (2011). Using conceptual spaces to model the dynamics of empirical theories. In E. Olsson & S. Enqvist (Eds.), *Belief revision meets philosophy of science* (pp. 137–153). Berlin: Springer.
Gärdenfors, P., & Zenker, F. (2013). Theory change as dimensional change: Conceptual spaces applied to the dynamics of empirical theories. *Synthese, 190*, 1039–1058.
Hampton, J. (2007). Typicality, graded membership, and vagueness. *Cognitive Science, 31*, 355–384.
Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*, 192–232.

[6]We are greatly indebted to Frank Zenker and two anonymous referees for valuable comments on a previous version of this paper.

Chapter 12
Specification of the Unified Conceptual Space, for Purposes of Empirical Investigation

Joel Parthemore

Abstract Recent years have seen a number of competing theories of concepts within philosophy of mind, supplanting the classical definitionist and imagist accounts: among them, Jerry Fodor's *Informational Atomism Theory*, Jesse Prinz's *Proxytypes Theory*, and Peter Gärdenfors' *Conceptual Spaces Theory* (CST). On the whole there has been little empirical investigation into the competing theories' merits; the (limited) empirical investigation of CST offers the one obvious exception. Some theories, such as Informational Atomism, seem almost beyond the possibility of such testing by design. Some philosophers would claim that theories of concepts, by their nature, cannot be tested empirically; and they raise valid concerns. Although I concede that theories of concepts are not open to direct empirical investigation, nonetheless indirect methods can provide strong circumstantial evidence for or against a theory such as CST; and I offer a research plan for doing so. Indeed, I argue that an extension of CST I call the Unified Conceptual Space Theory (UCST) is better placed than the competition when it comes to such testing, not least because it comes with a software application, in the form of a mind-mapping program, as a more-or-less direct translation of the theory into a working computer model. This paper provides the most detailed specification to date of the algorithm underlying the UCST, described in earlier publications as an attempt to move CST in a more algorithmically amenable and therefore, it is hoped, more empirically testable direction. UCST brings all the many widely divergent conceptual spaces discussed in CST together into a single unified "space of spaces" arranged along three axes, where points in the space have both local and distal connections to other points.

J. Parthemore (✉)
Centre for Cognitive Semiotics, Lund University, Lund, Sweden
e-mail: joel.parthemore@semiotik.lu.se

F. Zenker, P. Gärdenfors (eds.), *Applications of Conceptual Spaces*,
Synthese Library 359, DOI 10.1007/978-3-319-15021-5_12

12.1 Introduction

"Concepts", as I use the term, can be understood in two complementary ways:

1. The "building blocks" of structured thought.[1]
2. The abilities by which one thinks in a structured fashion.

That is, concepts are both things we possess and employ, and things that we *do*. Regardless of which of these two perspectives one takes, concepts are typically described as systematic, productive, and compositional (see e.g. Evans 1982, pp. 100–104 or Prinz 2004, pp. 12–14); attaching to an agent while mostly concerning things that are not the agent (*intentional*, in the Brentano sense,[2] or, as it is also commonly expressed, *referential*[3]); under the agent's *endogenous control* (Prinz 2004, p. 197); and, on most accounts other than Jerry Fodor's (1998) *Informational Atomism Theory*, subject to revision (if not, as on some accounts – e.g., Parthemore (2013) – in a state of constant if incremental motion, not too flexible so as to lose their systematicity but *flexible enough* to adapt to fit each new context of application). *Contra* certain accounts, so long as they meet the above conditions, I will not take concepts to be necessarily articulable (see e.g. Newen and Bartels 2007; Allen 1999) or even necessarily introspectible: i.e., using Gilbert Ryle's (1949) distinction, they may sometimes be more toward the *knowing how* than the *knowing that* side of the ledger.

On most accounts of the history of the field that has come to be known as *theories of concepts* within philosophy of mind – e.g., Prinz (2004), Fodor (1998), Margolis and Laurence (1999) – the story begins with so-called *classical definitionism*, whereby concepts are like entries in a dictionary, establishing necessary and sufficient conditions for their proper application. Jesse Prinz (2004) gives equal billing to *classical imagism*, whereby concepts are meant to be or be like "pictures in the mind" – an obvious proponent of which would be someone like George Berkeley (1999). Both traditions are now long out of fashion – classical definitionism because, as Fodor puts it (1998, p. 45), "there are practically no defensible examples of definitions"; classical imagism because of its reliance on

[1] As Sect. 12.3.2 points out, this metaphor, however helpful, breaks down fairly quickly.

[2] "Every mental phenomenon includes something as object within itself, although they do not do so in the same way. In presentation, something is presented, in judgment something is affirmed or denied, in love loved, in hate hated, in desire desired and so on" (Brentano 1995, p. 88).

[3] A theory of reference is really outside the scope of this paper, although the theory of concepts discussed in this paper owes no small debt to e.g. Saul Kripke's (1980) *causal theory of reference*, as one of the anonymous reviewers of this article pointed out. Nonetheless, it departs from the canonical causal theory in a number of ways, not least in its interest in conceptual reference as something at least partly distinct from linguistic reference.

resemblance as an *explanans* rather than an *explanandum* – the usual reference here being Nelson Goodman (1976, p. 5); both because of their reliance on a problematic notion of *representation* (symbolic and iconic, respectively) (Harvey 1992): for many researchers, including myself, the common-variety notion of *representation* inevitably raises the spectre of a homuncular view of cognition. (Who is doing the representing, and who is being represented *to*?).

Recent times have seen the appearance of, among others, prototype and exemplar theories, beginning with Eleanor Rosch (1999, 1975); Fodor's *Informational Atomism Theory*; Prinz's *Proxytypes Theory*, which may be taken as informational semantics without the atomism (Prinz 2004, p. 164; see also Parthemore 2011a, pp. 23–24); and Peter Gärdenfors' (2004) *Conceptual Spaces Theory* (CST). Fodor's account is in the rationalist tradition; Prinz's and Gärdenfors' accounts are in the empiricist tradition and owe an explicit debt to prototype theory.

For all that the nature of concepts is a matter of lively contemporary philosophical debate, remarkably little empirical research has been done. Prototype theory, with its talk of basic level, subordinate, and superordinate categories, has had a fair amount of empirical testing but is somewhat out of fashion, unless substantially modified *per* e.g. Prinz or Gärdenfors. As a product of the rationalist tradition, Informational Atomism Theory is almost designed to be beyond possibility of empirical testing: Fodor takes it (or at least something very much like it) to be logically necessary (which perhaps it is, if one assumes Fodor's metaphysics: nothing more can be said about the relationship between the concept of *dog* and dogs than that the concept reliably – under normal operating conditions – tracks all and only dogs). It is unclear how one would go about testing Proxytypes Theory, nor has Prinz offered any account of how one might do so. Of the above accounts, only CST has faced any kind of empirical investigation. As an extension of CST explicitly informed by enactive philosophy (see below), the *Unified Conceptual Space Theory* (UCST) (Parthemore 2013, 2011a; Parthemore and Morse 2010) is designed from the ground up to be amenable to empirical investigation.

12.2 Empirical Investigations Into (Theories of) Concepts

12.2.1 Why Conceptual Spaces Theory*?*

Setting issues of empirical testing aside for a moment, CST, couched in the relatively neutral language of geometry, provides a way to break out of thinking of concepts as "symbols in the brain" *per* Fodor (1998) or, from a somewhat different angle, Alan Newell (1980) – or, for that matter, classical so-called *cognitivist* cognitive science and AI in general. On the CST account, concepts sit *below* the level of symbolic thought (much if not most of the time conscious and deliberate, and indisputably *personal*, to use Daniel Dennett's (1969) *subpersonal/personal* distinction) but *above* that of (purely) association-based thought (largely unconscious

and automatic, while shading over into the subpersonal). Concepts are beholden to neither one level nor the other (Gärdenfors 2004, pp. 1–2). They are, in their most raw form, partitions in an *n*-dimensional Voronoi tessellation.

CST keeps what I think are the best parts of prototype theory, putting Rosch's notion of the *basic level of concepts* into a larger framework[4] while explaining, via its explicit emphasis on convexity as a standard property of concepts, why some concepts, because they are *not* convex (e.g., *Gentile = non-Jew*), do not have prototypes, as well as why some prototypes have larger "footprints" than others – something that the original prototype theory did not address. It employs a notion of similarity that does not fall foul of Nelson Goodman's dictum: similarity is *explanandum* rather than *explans*, determined by one's choice of integral dimensions for a given conceptual space and by the pre-determined metric for that space. To wit, the closer two points in a conceptual space are within that space, the more similar they are judged to be; the further apart, the more dissimilar. CST is unabashedly representationalist, but in a way that someone like Inman Harvey (1992) – widely regarded as an anti-representationalist (though he himself would dispute that claim [personal communication]) – could, I think, be comfortable with. Representations can be understood as inherent in the perspective of the *observer* of conceptual spaces and not (or not necessarily; I find Gärdenfors to be unfortunately somewhat unclear on this point) in the concepts or conceptual spaces themselves. Finally, with its tendency toward conceptual constructivism and metaphysical antirealism – conceptual spaces are meant to be something close to a self-supporting structure (think of Donald Davidson's (1986) *coherentism*), with no definitive claims about the pre-conceptual world – and with its affinity to a view on cognition that is simultaneously *situated*, *embodied*, and *extended*, CST is at least implicitly enactive, as Gärdenfors acknowledges (personal communication).[5]

CST is quite explicit about not trying to offer anything like a complete account of concepts but rather a scaffolding for others to build on. Gärdenfors writes:

> Philosophers will complain that my arguments are weak; psychologists will point to a wealth of evidence about concept formation that I have not accounted for; linguistics [sic]

[4] "There tends to be a privileged way of clustering the objects… that will generate the *basic categories* in the sense of prototype theory.… This is the set of clusters that provides the most 'economic' way of partitioning the world" (Gärdenfors 2004, p. 194).

[5] Note, too, that, as I write in (2011b, p. 86), "anti-realism, pragmatism, and pluralism go hand in hand, where pragmatism/pluralism is taken as the position that there need be, in most instances at least (and perhaps all of import), no single fact of the matter. Pragmatism can tolerate apparent contradictions, so long as they are qualified by perspective: e.g., *p* from one perspective, ~*p* from another. So long as one does not try to hold both perspectives at once – i.e., make them part of a single unified perspective – there is no contradiction in defending both." As this, perhaps, implies, antirealism can be more of a metaphysical position; or it can be more of a pragmatic one, in the spirit of American pragmatism, in which case it is metaphysically neutral: see William James' "fourth misunderstanding" of pragmatism (1975 [1909], p. 104).

> will indict me for glossing over the intricacies of language in my analysis of semantics; and computer scientists will ridicule me for not developing algorithms for the various processes that I describe. I plead guilty to all four charges (Gärdenfors 2004, p. ix).

12.2.2 Why Empirical Investigations?

Philosophy has often, perhaps rightly, been pilloried for its tendency to resort to "armchair theorizing", often with no apparent practical applications nor possibilities for empirical investigation. At the same time, philosophers often see their role as holding empirical researchers to account, pointing out that, while the empirical researchers' methodologies may indeed be – if not generally are – sound, their results are often subject to what are, philosophically speaking, extremely dubious claims. Too wit – they say – most observations are open to multiple, partly overlapping and partly mutually exclusive, interpretations. There seems to be a tendency to slide off toward either extreme – "armchair theorizing" or uncritical observation – whereas, to those of us who are interested in so-called *experimental philosophy*, the most exciting work lies where the tension between the two is actively maintained in a way that keeps both the philosophy and the "hard science" honest.

Getting one's theory of concepts "right" – at least with respect to the area of application; one way to read Edouard Machery (2009) is that there may be no such thing as a universal theory of concepts – is arguably essential to understanding the relationship between concepts as the "building blocks" of structured thought, language (assuming, as the so-called *animal concepts* philosophers do (Newen and Bartels 2007; Allen 1999), that concepts and language pull apart), signs, symbols, and representations: the stock-in-trade of psychologists, cognitive linguists, semioticians, and so on. From the perspective of *enactive philosophy* (see e.g. Thompson 2007; Varela et al. 1991; Maturana and Varela 1992) – with its openness to "extended mind" (Clark and Chalmers 1998; Clark 2008) and its emphasis on the interaction of agent and environment, "bringing each other forth" (to paraphrase Maturana and Varela 1992, p. 255) in such a way that the interaction is irreducible – it is essential as well to breaking down what enactivists see as the artificial barriers between "inner" thought and "outer" experience or communicative expression.[6]

[6]In this way, an enactive perspective acknowledges but goes beyond Hilary Putnam's externalism and Clark and Chalmers' contrastive *active externalism* to stake out a much more radical position whereby agent is ultimately continuous with environment, and both internalist and externalist perspectives on cognition are ultimately mistaken. Meanwhile, for more on why the self/other or self/world distinction should be taken as a conceptual rather than a prior ontological one, see Parthemore (2011b).

12.2.3 Why Not?

Although we can, and do, talk about our concepts – any reasonably well-educated person will understand and be able to at least attempt to answer a question like "explain to me, what is your concept of *happiness*?" – no one, it seems, has ever *seen* (or touched or otherwise observed) a concept. Given the current fashionableness of neuroscience, one might look for *neural correlates of concepts* as others are looking for *neural correlates of consciousness* (see e.g. Metzinger 2000). Unless one assumes some reductionist-type approach, however, concepts – as opposed to their neural correlates (if such even exist) – appear to be things that one cannot observe or measure directly. Indeed, the lack of empirical testing of theories of concepts might be taken to imply that the very idea of such investigations is highly controversial.

Besides the lack of observability, one explanation might be that concepts (as opposed to *concepts of*...) are too fundamental to everything, so that any discussion of them quickly becomes a *meta*-discussion. Indeed, a theory of concepts, by its nature, implicitly is such a meta-discussion already, presenting as it does not one or another person's "actual" concepts but the given researcher's concept of what those concepts are. A theory of concepts is a conceptual entity whose target is other concepts. Like any conceptual entity, a theory of concepts is *meant to be* distinguishable from the thing it refers to, even while it may (perhaps naively) be taken to refer to that thing transparently. As I wrote in Parthemore (2011a, p. 91):

> If concepts are, by their nature, necessary fictions, then any theory of concepts, as itself a conceptual entity, can be no more. If concepts simplify and, by their simplification, distort the reality they simplify away from, then so will a theory of concepts. If concepts depend upon some essential inconsistency between what they represent and what they purport to represent, then a theory of concepts will be similarly dependent.

The problem can be expressed this way: whatever approach is taken, the theory is being put forward from within a pre-existing conceptual structure, which it then purports to uphold. The broader the theory attempts to be, the greater the risk. Logically speaking, one cannot have a part (the theory of concepts in this case) successfully capturing the whole, which a too-broad theory of concepts must attempt to do. It's like the dragon swallowing its own tail.

Finally, putting all of this another way, one might argue – as I think Fodor is implicitly inclined to – that theories of concepts belong in the realm of metaphysics; and metaphysics is, almost by definition, not empirically testable. Rather, metaphysics provides a set of starting assumptions, like axioms in mathematics: *if* we assume these things, *this* is what follows. The proof of one or another set of metaphysical assumptions lies in their subsequent usefulness rather than any customary notion of truth.

These concerns are valid – *if* one tries to push one's theory of concepts too far; better to try to delimit one's theory to a particular field (say, a philosophical theory that should interest the psychologists, rather than a theory for all researchers everywhere) and a fairly well defined area of application (say, computer models of mind and learning support tools). Too, there seems to be good reason to think

that *direct* empirical investigation, with its requirement on directly observable and measurable properties, is impossible.

Assuming that's true, just because theories of concepts are not amenable to *direct* empirical testing does not mean that they cannot be tested in a variety of extremely productive *indirect* ways. An accumulation of circumstantial evidence weighs validly, if not decisively, for or against a theory such as CST or UCST that permits it. The more algorithmically well defined, the easier the theory is to test with, say, computer modeling; the more direct the translation from theory to implementation, the greater the likelihood that the implementation will reveal both strengths and weaknesses or gaps in the theory. That is a significant part of the driving motivation behind the development of UCST. First, however, I need to say something more about CST and its empirical testing to date.

12.2.4 What's Been Done So Far?

In two papers, Antonio Chella (Chella et al. 2008, 2004) offers what might be considered a partial implementation of CST on a robotics platform, via a set of "hard-wired" (as opposed to dynamically evolved) conceptual spaces. The focus of both papers is less on empirical results than on applying CST to the problem of *perceptual anchoring*, defined as "the problem of creating and maintaining in time the connection between symbols and sensor data that refer to the same physical objects" (Chella et al. 2004, p. 40), taken to be a special case of Stevan Harnad's (1990) symbol-grounding problem. A recent analysis of colour terms from 110 languages collected via the World Colour Survey appears to confirm the empirically testable claim from CST that colour terms should be universally convex, in keeping with the convexity of most concepts (Jaeger 2010).[7] Work on action spaces and force vectors is ongoing (Gärdenfors and Warglien 2012; Warglien et al. 2012; Gärdenfors 2007), though, with the exception of the empirical studies reported in Hemeren (2008) regarding action spaces, along with re-interpretation of earlier results with e.g. the *patch light* paradigm (Johansson 1973), that work remains largely theoretical, setting out possibilities for future empirical work just as this paper is trying to do. What remains largely missing is any testing of CST as such; implementations, such as Chella's, or re-interpretation of existing results, are only a step in the right direction.

[7]This is, perhaps, not so surprising, given that convexity is an implied property of most if not all prototype-based theories of concepts; the contribution of CST lies in making that commitment explicit and providing a geometry-rich account in which to understand it.

12.3 The *Unified Conceptual Space Theory* (UCST)

In contrast to CST, UCST is explicitly enactive: concepts are located *not* in the agent *nor* in the environment but are rather to be found in the interaction of agent and environment. Likewise, UCST makes explicit CST's implicit commitment to concepts being simultaneously "private" and "public," both idiosyncratically individual and irreducibly social.[8] According to UCST, and in keeping with lessons from the "animal concepts" philosophers, concepts exist and can be described at multiple levels:

- That of the individual agent ("my" concept of dog or "your" concept of dog) – perhaps the only available level for non-linguistic agents.
- That of the group.
- That of the society.
- That of the society as lexicalized into the words of a language.

UCST was designed from the ground up to fill some of the gaps in CST whilst fleshing it out with a specific algorithm: a kind of general "recipe" for concepts, be they abstract or concrete; objects, actions, or properties; basic-, superordinate-, or subordinate-level; types or tokens. To the extent they conform with the theoretical framework, all concepts should, so CST strongly implies, take the same basic structure: an empirically testable claim, born out by what preliminary testing has been done to date (see e.g. Hemeren 2008 on comparing "object" with action concepts). Such an approach comes with obvious benefits and drawbacks: on the one hand, a well-defined algorithm is easier to test empirically; on the other, it necessarily excludes any aspects of the target that do not, for reasons of practicality or principle, fit neatly into algorithmic form. There is no reason to think that concepts should be any different this way than any one or another conceptual domain one might wish to describe algorithmically. A theory of concepts is a (semi-)formalized system necessarily embedded in the seemingly much less formal structure of our concepts and conceptual frameworks more broadly – not unlike the way that strictly formal systems are necessarily embedded in the much less formal natural-language ones within which they are discussed and debated.[9]

There is the risk, too, that one inevitably over-intellectualizes matters, by taking a very *conceptual* view on concepts: something of how concepts get possessed and employed non-reflectively – as they must most of the time do – gets lost. At the same time, as Sect. 12.2.3 made clear, one cannot set one's conceptual nature aside to view concepts objectively and dispassionately; to a not inconsiderable extent, a conceptual view on concepts is all one can ever have.

[8] This amounts to a relaxation of the so-called *publicity principle* (Fodor 1998, p. 28) (but see also Prinz 2004, pp. 158–159). As one of the anonymous reviewers noted, this might usefully be compared to de Saussure's (2013 [1916]) distinction between *langue* and *parole*.

[9] H.B. Curry (1963) made precisely this point in distinguishing between *U-languages* ("usual") and *A-languages* ("artificial") – with thanks to one of my reviewers for pointing this out.

With that caveat in mind, UCST offers a "just so" story of how all concepts describable within its framework can be derived from three primary protoconceptual entities: proto-objects (a more stable *something*, corresponding roughly to the English grammatical category of "noun"), proto-action/events (a more dynamic *something*, corresponding roughly to the grammatical category of "verb"), and proto-properties (corresponding roughly to the grammatical categories "adjective" and "abverb"). All concepts trace back to one of these three proto-categories, which represent an initial, minimal, innate[10] partitioning of the unified space (Parthemore 2014).

Certain further protoconcepts are required: two requisite proto-objects are *proto-time*, which can be described along two dimensions, one from past to future and one of alternative events ("might-have-beens"); and *proto-space*, which can be defined along three dimensions. Proto-space further divides into *proto-physical-space* and *proto-conceptual-space*. One requisite proto-property is *proto-dynamic* (see below).

UCST attempts to describe how all of an individual conceptual agent's conceptual spaces – and, by extension, all the conceptual spaces of all the conceptual agents in a given society – map or weave together into a single, unified space: a multidimensional "space of spaces" describable along three common axes defining the proximal connections of points in the unified space. As such, it makes certain empirically testable claims beyond those made already by CST:

- All concepts, as concepts are defined by the theory, can be adequately described within this schema. Although UCST does *not* attempt to be a universal theory of concepts (in that it is not intended to be applicable to *all* possible areas), nevertheless, if true, this does give concepts psychological reality and so contradict the claim made by Edouard Machery that "progress in the psychology of concepts... is conditional on psychologists... eliminating the notion of concept from their theoretical vocabulary" (Machery 2009, p. 4).
- Furthermore, all concepts can be adequately described as falling within one of the three proto-categories of proto-object, proto-action/event, and proto-property.[11]
- Supporting evidence for the innateness of these protoconcepts should be forthcoming from child development studies, particularly those looking at newborn and very young infants.

[10]How one cashes out "innate" lies beyond the remit of this paper. What matters is that nativism is a severely restricted nativism, as opposed to something like Fodor's *radical nativism*. Note, too, that on Fodor's (1998) account, the relationship between essentially *all* concepts and their referents is (what he terms) nomic; according to UCST, the only nomic relations apply to protoconcepts.

[11]This has the consequence, of course, that not all words of a language express (lexical) concepts, but that is a consequence that UCST, following in the tradition of the "animal concepts" philosophers who claim that concepts and language pull apart, is happy to embrace.

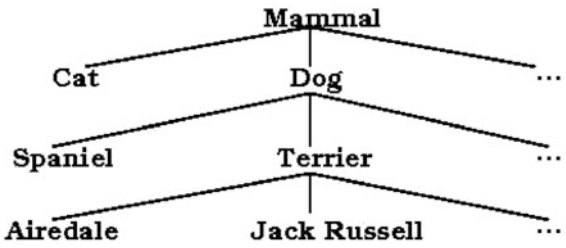

Fig. 12.1 The ISA hierarchy (Reproduced with permission from Hemeren 2008, p. 25)

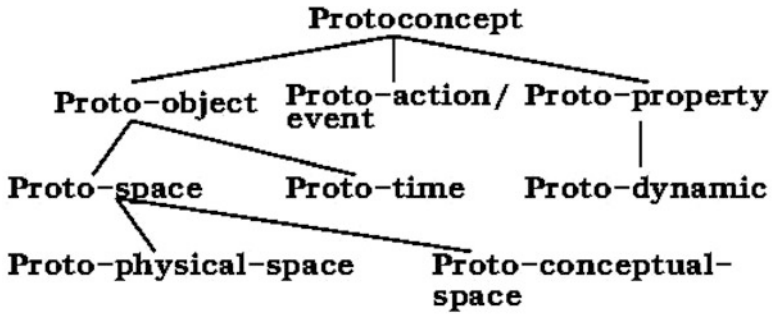

Fig. 12.2 The minimal ISA hierarchy in the unified space

12.3.1 Proximal Connections

12.3.1.1 Axis of Generalization

This reprises the familiar ISA hierarchy from maximally general to maximally specific, whereby e.g. an Airedale is a terrier is a dog is a mammal is an animal… (see Fig. 12.1, where each possible path reflects a different trace along this axis: e.g., *Airedale* → *terrier* → *dog* → *mammal* or *spaniel* → *dog* → *mammal*). At one end, everything is a *something*. At the other, everything is a *particular*. The minimal requisite structure distinguishes *proto-object* from *proto-action/event* from *proto-property* (see Fig. 12.2 for further detail). New nodes can be added according to either of two rules:

1. Given whatever is currently the maximally specific node on a particular trace along the axis, add an additional node via a single ISA link until one no longer has a class but a solitary instance of the superordinate class. This comes with the caveat that every instance is – looked at the right way – a potential class or collection by some alternate description (e.g., a human being can be viewed as a collection of specialized cells). In other words, there is no principled class/instance distinction, a point I take (Parthemore 2011a, p. 117) to be implicit in CST and which is made explicit in the UCST.
2. Given any two nodes on a particular trace, beyond the initial two protoconceptual nodes – the protoconceptual structure cannot be changed – add an additional

node via two ISA links: e.g., one might have *Kali* ISA *mammal* and change this to *Kali* ISA *cat* ISA *mammal*. Given the opportunity to add arbitrarily many such nodes, the axis of generalization should be understood as a continuous line, certain points along which are marked out with labels.

12.3.1.2 Axis of Alternatives

As Fig. 12.1 suggests, at any point on a particular trace along the axis of generalization, one can "take a right turn" to explore alternative (XOR) options: a *dog* is a *mammal*, but so is a *cat* (but a *dog* cannot be a *cat*); a *terrier* is a dog, but so is a *collie*. Likewise, *blue* is a colour, but so is *green*. These alternatives lie along what I have called the *axis of alternatives*, derived by incrementing or decrementing the value of any one or more of the integral dimensions defining that concept, according to the predefined metric of the space. Depending on the dimensions in question, these could be numerical over a finite, cyclical, or unbounded (infinite) range (quantitative); or they could be merely ordinal, positioned from most to least similar according to some further metric for similarity (qualitative). Any possible combination of integral dimensions constitutes a particular trace along the axis of alternatives.

To take Gärdenfors' (2004) most commonly used example: *colour* has the integral dimensions of *hue*, *saturation*, and *brightness*, whose predefined metric could simply be a numerical value between 0 (none, in the case of *saturation* or *brightness*) and 255 (full). In this way, a single point along the axis of generalization opens up into a space; what was the point *colour* becomes e.g. the colour cone or sphere, depending on how it is mapped out (see Fig. 12.3). Indeed, along this axis one finds all of the conceptual spaces that Gärdenfors talks about in (2004).

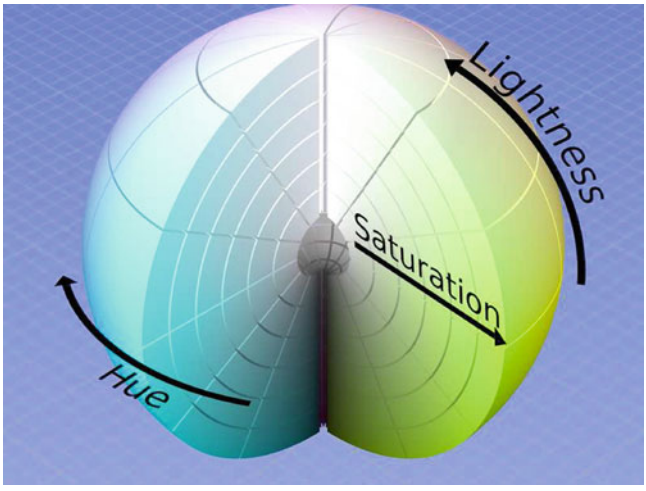

Fig. 12.3 The colour space (Reproduced from Wikimedia Commons: http://commons.wikimedia.org)

The initial minimal structure distinguishes *proto-object* from *proto-action/event*, according to the integral dimension of *dynamics*. ([*Proto-*]*objects* are less dynamic, [*proto-*]*action/events* more so.) Once again, new nodes can be added via either of two rules, at any point along the *axis of generalization* below the protoconceptual level.

1. Given whatever is the current terminal node (assuming a non-cyclical dimension), add an additional node via a single XOR link: e.g., in Fig. 12.1, *bulldog* might be added beyond *terrier* on the judgment that a *bulldog* is less like a *spaniel* than a *terrier*.
2. Given any two nodes, add an additional node via two XOR links: e.g., if one has *spaniel* and *poodle*, one might add *cavapoo* between them. If one has *yellow* and *blue* along a particular trace through the colour space based primarily on *hue*, one could add either *green* between *yellow* and *blue* or *red* between *blue* and *yellow*, since the hue dimension is cyclical.

12.3.1.3 Axis of Abstraction

On any particular trace along the axis of generalization, another kind of "right turn" is possible, exploring a different kind of alternative (IOR) options and a different kind of abstractness. Points along this axis are arranged from most physical/concrete (least obviously conceptual or even seemingly fully non-conceptual) to most mental/abstract (explicitly or higher-order conceptual; see Fig. 12.4).

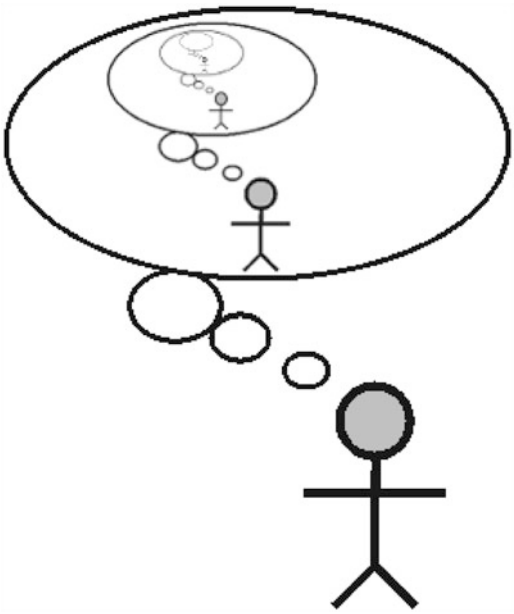

Fig. 12.4 From thoughts to thoughts about thoughts (Adapted from an image downloaded from Wikimedia Commons: http://commons.wikimedia.org)

Arranged into discrete intervals, one has a progression from "zeroth"-order concepts (non-concepts) to first-order concepts (concepts of non-concepts), second-order concepts (concepts of concepts), and higher-order concepts (concepts of concepts of concepts).[12] In the case of *colour*, movement along this axis implies a shift from *colours-as-experienced* (most concrete) to the *colour sphere/cone* (more abstract) to e.g. *colour-as-judged-aesthetically* (most abstract). In the case of cats,[13] one has a progression from *cat-as-experienced* (under the most fine-grained-possible spatiotemporal or conceptual description) to *cat-as-biological-entity* to *cat-as-possibly-selfconscious-intentional-agent* to such truly esoteric levels as *how-I-think-that-I-think-about-cats* (my concept of my concept of *cat*).

Importantly, all points along the axis of abstraction represent one and the same target. In this way, the familiar *sense* vs. *reference* distinction (Frege 1892) between *what* a concept is about and *how* it goes about being about it[14] collapses. In complete reversal to Fodor's program, whereby sense ultimately collapses into reference (Fodor 2008, p. 151), here, reference ultimately collapses into sense. Reference, in this system, is simply a mapping from a point "higher" on the axis to an arbitrary point "lower" on it.[15]

The initial minimal structure distinguishes *proto-property* (more abstract) from either *proto-object* or *proto-action/event* (more concrete). The same basic two rules apply:

1. Whatever is the current terminal node (maximally concrete or maximally abstract), add a new terminal node via a new IOR link, up to the practical limits: beyond some point, one can't be more precise about spatial location or temporal occurrence; likewise, beyond some point, one loses oneself in conceptual abstractness (a concept of a concept of a concept of...).

[12]It is, indeed, very possible that this axis and the axis of generalization meet at the one extreme, so that a maximally general concept and a maximally abstract or higher-order one amount to the same thing – a possibility I explicitly allow in Parthemore (2011a, p. 130). Note that the converse is *not* the case: a maximally specific concept and a maximally concrete or lower-order one are *not* the same thing!

[13]... and other mammals, and perhaps other species – depending on how far one chooses to push (self-)consciousness.

[14]The concern, of course, is that one must account for conceptual content that cannot be provided solely by reference. One can possess the concept of the *Morning Star* and the concept of the *Evening Star* without realizing they are co-extensive.

[15]To clarify the relation between Frege's position, Fodor's, and mine: on Frege's account, *unicorn* has a sense (it refers in a certain sort of way) but not a reference (it does not, in fact, refer *to* anything). Fodor takes a counterfactual approach: to possess the concept *unicorn* entails that, were there any unicorns, the concept would lock to (all and only) unicorns. On my account, *unicorn* does have a referent, just of a different kind from e.g. *dog*, just as *tranquility* also has a different kind of referent from *dog*. So far as anyone knows – surgically altered circus animals aside – unicorns exist only in mythical worlds, the stuff of imagination, of fairy tales and fiction; dogs exist there, too, but also exist in the physical world. The difficulty, if any, with *unicorn* is that it appears to have the same sort of referent as *dog* and it does not: one is a mammal; one is something else.

2. For any existing two adjacent nodes, add an additional node between them via two IOR links (e.g., *cat-under-an-anatomical-description* may be judged more abstract than *cat-as-biological-entity* but more concrete than *cat-as-possibly-selfconscious-intentional-agent*).

12.3.2 Distal Connections

Within the UCST framework, all concepts and protoconcepts map to distal parts of the unified space in two ways: all may be described by certain *parameters*; and all exist within the context of other, frequently or occasionally associated *contextual* elements: i.e., things that their target is commonly co-present with. In addition, certain concepts describe mereological (part/whole) *component* relations.

(This is one of the reasons why the concepts-as-building-blocks metaphor breaks down. With children's building blocks, the important relations are strictly proximal: one block going above, below, or beside another block. Distal connections among ordinary building blocks would suggest something "spooky"; if one moves one block, a distant block will not move unless all the blocks in between move as well, in domino-like effect.)

12.3.2.1 Components (Mereological Relations)

Object concepts and action/event concepts, more toward the concrete than the abstract end of the axis of abstraction, decompose into parts: objects into objects, action/events into action/events, like into like. (Abstract object and action/event concepts generally do not decompose in this way, and property concepts are even less likely to: one does not typically find a concept that consists e.g. of being green and being heavy, or being young and being talented. In general, the more abstract a concept is, the less likely it is to decompose.) These must be ordered: i.e., their order of composition cannot be arbitrary; and one or more of them must be necessary, while the remainder may be optional. Components inherit: *live person* has the necessary component *head*, because *animal* (or perhaps a subset of *animal*) has the necessary component *head*; *live person* has the additional necessary component *torso*, which must be in a certain ordered relation to *head*; while *arm* and *leg* are, as it were, optional: one can lose one's arms and legs, or be born without them, but still qualify as a living person, but a *person* with no *head* or no *torso* is something else, albeit person-related: e.g., a corpse.[16] In similar fashion, *to pitch* has the necessary

[16]What precisely the necessary part(s) is/are is not important: these can at least in part be pragmatically determined by present context and subject to change. If one meets a living person who *is* just a head and nothing more – attached, say, to a prosthetic robot body – one might well revise one's *living person* concept to incorporate this new possibility.

components *to grasp* and *to release*, inheriting these from *to throw* (noting that *to grasp* must come before *to release*). *To pitch* adds the additional necessary component of *to take aim*; one can throw something aimlessly, but if one pitches something, either metaphorically or "literally", one necessarily has a goal or target in mind.

12.3.2.2 Parameters (Integral Dimensions)

What I am calling *parameters* constitute *integral* (individually and jointly necessary) as opposed to *separable* dimensions[17]; at the same time, they define the *axis of alternatives* (Sect. 12.3.1.2) and constitute different conceptual spaces or domains within the unified space: the *colour* space, the *tone* space, etc. They are not ordered: e.g., *colour* has the integral dimensions *hue, saturation*, and *brightness*, but there is no ordering to be made among the three: *hue*, *saturation*, and *brightness* is precisely the same as *saturation, brightness*, and *hue*. Likewise *physical object* has the integral dimensions of, say, *weight*, *mass*, *volume*, *colour*, and *transparency*, while *activity* has at least the integral dimension of *duration* and perhaps also of *purpose*. A concept inherits parameters from its superordinate concept(s) but must in some way further specify or add to these so as to distinguish itself from the superordinate concept(s). Just as components are always either object or action/event concepts, parameters are always property concepts.

12.3.2.3 Contextuals (Figure/Ground Relations)

No concept can be given coherent interpretation except with relation to some non-empty set of contextual elements, which collectively define its contexts of encounter and application: in slogan form, no concept exists in a vacuum. Henceforth, I will refer to these simply as *contextuals*. Contextuals are neither ordered, like components, nor necessary, like parameters or at least a subset of any given components. They can be weighed only by their relative likelihood. Like components and parameters, they inherit; however, that inheritance can be overridden: *bird* has the contextual element *to fly*, but *ostrich* or *dodo bird* does not. Putting this another way, because the contextual relation is customary rather than necessary, contextual inheritance is always *ceteris paribus*. Although the presence of any one or another contextual element is always optional, specification of a sufficient number of them may limit the possible scope of a particular concept to a specific instance of its application.

The contextuals of object concepts split into two groups: action/event concepts and other object concepts. On the one hand, an object shares a certain (physical or conceptual) space with other objects; on the other, it is associated with or involved in

[17]The terms are taken from the psychology literature. For a good introduction, see Gärdenfors (2004, pp. 24–26).

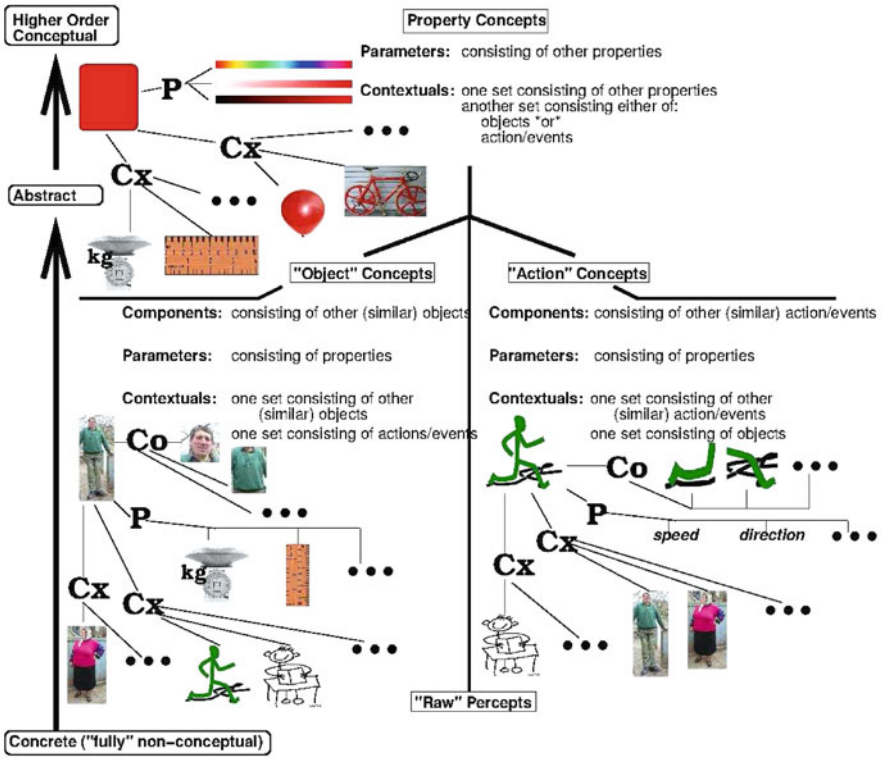

Fig. 12.5 Varieties of concepts in the unified space, and how they are structured: **Co** = components, **P** = parameters, **Cx** = contextuals

certain actions or events. The contextuals of action/event concepts split into the same two groups. Actions and events do not take place in isolation but in the context of other actions and events. They take place as well in a context of the agents initiating the actions and the entities with which the actions/events interact. Finally, the contextuals of property concepts split into two groups: on the one hand, one finds a set of other property concepts, consisting of *separable dimensions* (where one finds *colour*, one typically finds *mass* and *density* as well); on the other, *either* objects of which they are properties *or* action/events of which they describe the properties. No property concept attaches both to an object concept and to an action/event concept.

Figure 12.5 offers an illustration of the three main groupings of concepts in relation to the three types of distal connections and the axis of abstraction.

12.3.3 Software Implementation

An implementation of UCST exists as a direct translation of a slightly earlier form of the algorithm in the form of a *mind-mapping* program. A mind map is a paper- or

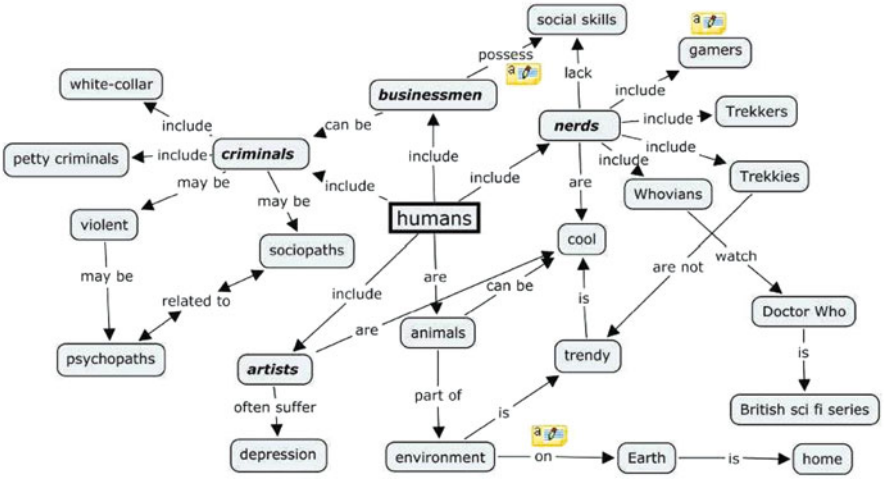

Fig. 12.6 A screenshot from CmapTools

computer-screen-based 2D visualization of a collection of related ideas and their interconnections, typically though not always arranged around a central idea (see Fig. 12.6). Mind-mapping software is often used for brainstorming ideas and for assisting students with certain learning disabilities. The idea is that, by not requiring ideas to be put down on paper or screen in a linear fashion, one mirrors the structure of non-linear subconscious thought and reduces cognitive load. As psychologist and neuroscientist Joseph LeDoux writes (LeDoux 1996, p. 280): "consciousness seems to do things serially, more or less one at a time, whereas the unconscious mind, being composed of many different systems, seems to work more or less in parallel". So, software tools and techniques that directly support that largely if not entirely unconscious level of cognition may assist e.g. the writing process in ways that tools aimed at the conscious/linear/propositional level may not.

For my purposes, the relevant aspect of mind mapping is the way it assists people in creating an "externalized" understanding of some portion of their conceptual domains. What is externalized is thereby made explicit; what is made explicit is easier to scrutinize; what is easier to scrutinize is *ceteris paribus* easier to modify.

One major shortcoming of currently available mind-mapping software is the lack of any well-defined theory of cognition, let alone theory of concepts, underlying the application – this despite the way that "mind maps have been invested with almost miraculous powers" (Sharples 1999, p. 80). Most particularly, these tools are, by any concept-theory-sensitive account, massively under-constrained, and there is no agreed technique on either how to construct or evaluate them. Mario Ruiz-Primo and Richard Shavelson write (1996, p. 585): "unfortunately we cannot look to cognitive theory to decide which technique to prefer, because many of the techniques [we] reviewed had no direct connection with such a theory".

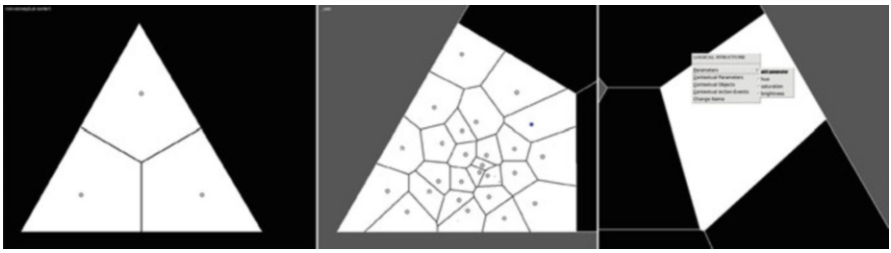

Fig. 12.7 Charley: initial and subsequent screenshots. The two dimensions shown represent the axis of alternatives; the axis of generalization runs vertically through the image. The *dots* both represent prototype instances of each concept and serve as gateways to "deeper" levels of the structure along the axis of generalization

The implementation of UCST – a program named Charley, created by the author for his doctoral thesis (2011a) – differs from existing mind mapping software in two important ways. First, it is informed by a specific theory of concepts; and, second, although it has the same functionality as existing software, visually it has a strikingly different interface, being based as are CST and UCST on Voronoi tessellations. Instead of creating nodes and establishing arbitrary links between them, one increasingly partitions an initially minimally partitioned space, diving down into that space and creating links between distal regions of it (see Fig. 12.7).

In its present form, Charley is a proof of concept. Currently, one can traverse freely only along the axis of generalization; the axis of alternatives is present but constrained (severely!) to the two dimensions portrayed in the screenshots; and the axis of abstraction is not present at all. For the axis of alternatives, it should be possible to designate any two parameters (integral dimensions) – presuming there are at least two – for visualization at any given time, and it should be easy to switch between which two one is viewing. For example, for the colour space, one should be able to choose any two of *hue*, *saturation*, and *brightness*. Certain domains – e.g., the *colour* and *tone* spaces – should be pre-populated with sensible "perceptual primitives" (i.e., portions of the *colour* and *tone* spaces so far as possible toward the "concrete" end of the axis of abstraction).

12.3.4 Empirical Investigation of UCST

In consultation with Jordan Zlatev (linguist) and Daniel Barratt (psychologist) of the Centre for Cognitive Semiotics, University of Lund, Sweden, work is in the planning stages for two pilot studies, the goal of which is to evaluate the readiness and appropriateness of the algorithm and software for empirical investigation of theories of concepts; more precisely: (1) a set of qualitative and quantitative measures benchmarking the UCST application against traditional mind-mapping software and equivalent pen-and-paper methods; (2) a psychologically valid measure of belief

consistency, determined by how frequently and in what ways users revise their mind maps over the course of the test session; and, most critically, (3) a list of errors and omissions in the program and the underlying theory.[18,19]

Remember that the software application is a direct translation of UCST theory. The goal is to explore how natural – or unnatural – the application appears to naive users,[20] as indirect evidence for or against some isomorphism, at some level of abstraction, between how it structures information and how their underlying conceptual thought processes do so: which is to say, some evidence that the theory is, in general and at least to some extent in detail, a psychologically plausible account of how concepts are structured. The proposed experiments do this by addressing such questions as:

- How much (if any) instruction do users require?
- How quickly (i.e., reaction times) and effectively do they complete their assigned task?
- How readable are their maps across users and by experts, compared to mind maps produced in more traditional ways?
- How do participants feel about the tasks on completion (as determined by nine-point Likert scales and free-form verbal reports)?

Less instruction, faster completion times, more "universally" readable maps, and more satisfied interaction all would lend indirect weight to the psychological plausibility of UCST – just as *more* instruction, *slower* completion times, *less* "universally" readable maps, and *less* satisfied interaction would all weigh in against it, at least in its current form.[21]

[18]The primary goal here, of course, is empirical testing of the software and underlying theory of concepts. That said, one of the inevitable outcomes will be to better adapt subsequent versions of the software to the requirements of the user, with the realization that there is a potentially commercially viable application here.

[19]As one of the anonymous reviewers noted, UCST might also be tested empirically using the techniques of *latent semantic analysis* on large corpora – a possibility I explicitly allow in Parthemore and Whitby (2013) and Parthemore (2011a, p. 190). That, however, I take to be dependent on further development of the theory and algorithm; whereas the empirical testing described here can – with some preliminary work required in re-implementing the software in a more efficient computer language with a more complete interface – be done now.

[20]Putting this another way: how familiar or strange do users find the software, particularly on first encountering it?

[21]One of the anonymous reviewers raised the question of whether the standard deviation on one or more of these measures might not, in fact, turn out to be quite high or the results fail to be reproducible. Although these things cannot safely be predicted in advance, nevertheless, if the experiments are well designed then one would, indeed, hope for a high degree of reproducibility. That is why attempts at reproducing results at other times and by other research teams are made: to discover whether the results *are* reproducible. Certainly high standard deviations are possible; their explanation would depend on follow-up experiments to discover what previously unconsidered variables might be causing them.

In both studies, participants will be presented with one of three possible applications: traditional mind-mapping program, UCST-based program, or pen-and-paper. Each participant will be given the task of creating three mind maps presented in random order: one of a noun-like "object" concept, one of a verb-like "action" concept, and one of an adjective-like "property" concept. Level of instruction will be varied systematically between participants: a first group will be given a standard level of instruction how to use the UCST application; a second group will be given minimal instruction: e.g., "do your best to figure out how this software works". The rationale for having two studies is to investigate whether any significant differences exist between how participants perform the task working individually (the first study) vs. how they perform it in collaboration with others (the second study, in which participants will work in self-selected small groups). In this way, degree of social interaction will be also manipulated.

Both studies will be based on and analyzed according to factorial design. The domain ("object" vs. "action" vs. "property") will be treated as a within-subjects factor; the type of application used (traditional vs. UCST-based vs. pen-and-paper), level of instruction (standard vs. minimal), and degree of social interaction (individual vs. group) will be treated as between-subjects factors.

In both studies, participants' overt behavioural responses will be recorded by video camera. Keyboard and mouse responses will be logged. In the first study only, participants' eye movements will be recorded by a remote eye-tracking system.

12.4 Conclusions

Not much empirical investigation of theories of concepts has been done to date, outside of certain, highly circumscribed areas. There are good reasons for this. Theories of concepts probably cannot be empirically tested directly – concepts are not directly measurable things – but indirect testing is possible and can be highly fruitful in ruling certain approaches or theories in or out.

Some theories of concepts, such as Fodor's Informational Atomism Theory, seem designed *not* to be empirically addressable; at the least, empirical investigation (as opposed to, say, logical analysis) is not the researchers' concern. Conceptual Spaces Theory tries to be empirically testable but, in a number of critical areas, lacks the required specificity. (Conceptual convexity, its strongest claim to proof-by-empirical-testing, is implicit in most if not all prototype-based theories of concepts.) UCST is not only designed from the ground up with an eye toward empirical testing – it makes a whole series of specific, testable claims – but it comes with a software implementation. By giving the software to naive users, it should be possible to see how they do – or do not! – get along with it.

Acknowledgements The author gratefully acknowledges the financial and academic support of the Centre for Cognitive Semiotics at Lund University, directed by Prof. Göran Sonesson and assisted by Prof. Jordan Zlatev; the assistance of Daniel Barratt in designing the experiments; and, finally, the helpful discussion and criticism received at seminars of the Centre for Cognitive Semiotics.

References

Allen, C. (1999). Animal concepts revisited: The use of self-monitoring as an empirical approach. *Erkenntnis, 51*(1), 33–40.

Berkeley, G. (1999). *Principles of human knowledge and three dialogues*. Oxford/New York: Oxford University Press

Brentano, F. (1995). *Psychology from an empirical standpoint*. London: Routledge.

Chella, A., Coradeschi, S., Frixione, M., & Saffioti, A. (2004). Perceptual anchoring via conceptual spaces. In *Proceedings of the AAAI-04 workshop on anchoring symbols to sensor data*, San Jose, California (pp. 40–45). AAAI Press, Menlo Park.

Chella, A., Frixione, M., & Gaglio, S. (2008). A cognitive architecture for robot self-consciousness. *Artificial Intelligence in Medicine, 44*(2), 147–154.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford: Oxford University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19.

Curry, H. B. (1963). *Foundations of mathematical logic*. Mineola/New York: Courier Dover Publications.

Davidson, D. (1986). A coherence theory of knowledge and truth. In E. LePore (Ed.), *Truth and interpretation* (pp. 307–319). Oxford: Blackwell.

Dennett, D. C. (1969). *Content and consciousness*. London: Routledge & K. Paul.

de Saussure, F. (2013 [1916]). *Course in general linguistics* (Trans. W. Baskin). New York: Columbia University Press.

Evans, G. (1982). *Varieties of reference*. Oxford: Clarendon Press.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford: Clarendon Press.

Fodor, J. A. (2008). *LOT 2: The language of thought revisited*. Oxford: Clarendon Press.

Frege, G. (1892). Über Sinn und Bedeutung. Zeitschrift für Philosophie und Philosophische Kritik C:25–50, original publication

Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*. Cambridge: Bradford Books.

Gärdenfors, P. (2007). Representing actions and functional properties in conceptual spaces. In T. Ziemke, J. Zlatev, & R. Frank (Eds.), *Body, language and mind, volume I: Embodiment* (Vol. 1, pp. 167–195). Berlin: Mouton de Gruyter.

Gärdenfors, P., & Warglien, M. (2012). Using conceptual spaces to model actions and events. *Journal of Semantics, 29*(4), 487–519. doi:10.1093/jos/ffs007

Goodman, N. (1976). *Languages of art: An approach to a theory of symbols*. Cambridge: Hackett Publishing.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena, 42*(3), 335–346.

Harvey, I. (1992). Untimed and misrepresented: Connectionism and the computer metaphor, cSRP 245.

Hemeren, P. (2008). Mind in action: Action representation and the perception of biological motion. PhD thesis, University of Lund.

Jaeger, G. (2010). Natural color categories are convex sets. In M. Aloni, H. Bastiaanse, T. de Jager, & K. Schulz (Eds.), *Logic, language and meaning* (Lecture notes in computer science, Vol. 6042/2010, pp. 11–20). Berlin/Heidelberg: Springer.

James, W. (1975 [1909]). *The meaning of truth*. Cambridge: Harvard University Press.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics, 14*, 201–211.
Kripke, S. A. (1980). *Naming and necessity*. Cambridge: Harvard University Press.
LeDoux, J. E. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon and Schuster.
Machery, E. (2009). *Doing without concepts*. Oxford/New York: Oxford University Press.
Margolis, E., & Laurence, S. (1999). Concepts and cognitive science. In E. Margolis & S. Laurence (Eds.), *Concepts: core readings*. Cambridge: MIT.
Maturana, H. R., & Varela, F. J. (1992). *The tree of knowledge: The biological roots of human understanding*. London: Shambhala.
Metzinger, T. (Ed.). (2000). *Neural correlates of consciousness: Empirical and conceptual questions*. Cambridge: MIT.
Newell, A. (1980). Physical symbol systems. *Cognitive Science, 4*(2), 135–183.
Newen, A., & Bartels, A. (2007). Animal minds and the possession of concepts. *Philosophical Psychology, 20*(3), 283–308.
Parthemore, J. (2011a). Concepts enacted: Confronting the obstacles and paradoxes inherent in pursuing a scientific understanding of the building blocks of human thought. PhD thesis, University of Sussex, Falmer, Brighton, UK. Available from http://sro.sussex.ac.uk/id/eprint/6954.
Parthemore, J. (2011b). Of boundaries and metaphysical starting points: Why the extended mind cannot be so lightly dismissed. *Teorema, 30*(2), 79–94.
Parthemore, J. (2013). The unified conceptual space theory: An enactive theory of concepts. *Adaptive Behavior, 21*, 168–177. doi:10.1177/1059712313482803.
Parthemore, J. (2014). Conceptual change and development on multiple time scales: From incremental evolution to origins. *Sign System Studies 42*, 193–218.
Parthemore, J., & Morse, A. F. (2010). Representations reclaimed: Accounting for the co-emergence of concepts and experience. *Pragmatics & Cognition, 18*(2), 273–312.
Parthemore, J., & Whitby, B. (2013). When is any agent a moral agent? Reflections on machine consciousness and moral agency. *International Journal of Machine Consciousness, 5*(2), 105–129.
Prinz, J. (2004). *Furnishing the mind: Concepts and their perceptual basis*. Cambridge: MIT.
Rosch, E. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573–605.
Rosch, E. (1999). Principles of categorization. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (chap. 8, pp. 189–206). Cambridge: MIT.
Ruiz-Primo, M. A., Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching, 33*(6), 569–600.
Ryle, G. (1949). *The concept of mind*. London: Penguin.
Sharples, M. (1999). *How we write: An account of writing as creative design*. London: Routledge.
Thompson, E. (2007). *Mind in life: Biology, phenomenology and the sciences of mind*. Cambridge: Harvard University Press.
Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge: MIT.
Warglien, M., Gärdenfors, P., & Westera, M. (2012). Event structure, conceptual spaces and the semantics of verbs. *Theoretical Linguistics, 38*(3–4), 159–193. doi:10.1515/tl-2012-0010.

Chapter 13
A Perspectivist Approach to Conceptual Spaces

Mauri Kaipainen and Antti Hautamäki

Abstract It is a part of everyday life that objects appear different from each perspective they are seen from. Ordinary language has plenty of expressions referring to abstract issues "from my point of view" or "your perspective". In this article, we argue for a *perspectivist* approach to conceptual spaces, that is, an approach to concepts as entities whose definition depends on the perspective from which they are considered. We propose an interpretation of Gärdenfors's conceptual space in terms of two components: a highly multi-dimensional *ontospace* whose simultaneous grasp is beyond or near the edge of human cognitive capabilities, and a lower-dimensional *representational* space that supports conceptualization of the ontospace in the manner Gärdenfors has suggested, however allowing several alternative conceptualizations, not just one. We suggest that a given ontospace is only accessible to the cognition by means of the epistemic work of exploring alternative perspectives. Further, we suggest that the overall understanding of a domain that emerges from seeing it from multiple perspectives is on a higher abstraction level than any particular single perspective. We stress that perspectives to the ontospace are individual and vary as a function of interest, situational contexts and various temporal factors. On the other hand, they are communicable, allowing interpersonally shared conceptualization.

13.1 Introduction

Gärdenfors's theory of conceptual spaces (1990, 1992, 2000, 2001) has made an significant impact on today's cognitive science, not only by means of providing a bridge between symbolic and connectionistic theories and semantics. Its influence

M. Kaipainen (✉)
School of Natural Sciences, Technology and Environmental Studies, Södertörn University, Alfred Nobels allé 7, S-141 89 Huddinge, Sweden
e-mail: mauri.kaipainen@sh.se

A. Hautamäki
The Department of Philosophy, History, Culture and Art Studies, Faculty of Arts, University of Helsinki, P.O. Box 59, FI-00014 Helsinki, Finland
e-mail: antti.hautamaki@kolumbus.fi

F. Zenker, P. Gärdenfors (eds.), *Applications of Conceptual Spaces*,
Synthese Library 359, DOI 10.1007/978-3-319-15021-5_13

on the development of the theories of categorization, induction and the emergence of language has been important, not least due to its contribution to the prototype theory (Rosch 1973, 1975). The assumption of similarity as the foundation of concepts and categorization that underlies Gärdenfors's work has a long preceding tradition in psychology and philosophy. One of the prominent theoreticians of similarity is Shepard (1987), who associates similarity to generalization. He also remarks that the issue of similarity is much older, even dating as far back as to Aristotle (ibid. 1317).

One of the main criticisms against similarity-based cognition worth bringing to the discussion is the idea that this similarity is too vague an idea to explain cognitive processes unless there is a definite account of what counts as a quality dimension (Murphy and Medin 1985; Gärdenfors 2000, 108). This issue is closely associated with the dynamic nature of conceptualization. Gärdenfors cites (ibid, 109) Goodman who pointed out that "similarity is relative and variable" (1978, 437). This criticism, in our view, does not, however, undermine the significance of similarity, but rather makes it compelling to analyze the factors with respect to which similarity is relative. While admitting that a full model of cognitive mechanisms should include the processes that operate on representations (Gärdenfors 2000, 31), he leaves such considerations outside of the model (in 2000 and 2001), but returns later to discuss them from various angles. A systematic inquiry of geometric representation of similarity is done by Johannesson (2002), who finds several new ways to the increase the descriptive powers of geometric models, one of them being the descriptive powers of geometric models can be increased in a number of ways.

Although the conceptual model has a large degree of explanatory value, the questions remain as to *when, how* and in *which context* the concept-constituting similarity occurs. The present paper aims to contribute to the analysis of the last two questions, and at least open the issue of the first. With *how* we point to the dynamical and interactive exploration of similarities, and *in which context* we refer to perspectives that determine similarity.

For Gärdenfors, a conceptual space is determined by *quality dimensions* of which some might be innate, some learned, or culturally dependent, and some even introduced by science (Gärdenfors 1992, 4). In his approach, concepts are regions of conceptual space augmented by the geometry and metric of the conceptual space. A key element appears to be that the knowledge representation is *non-linguistic* in the sense that "we can represent the qualities of objects without presuming an internal language in which these qualities are expressed". The qualitative dimensions are thereby ontologically prior to any form of language. This presupposes that it is possible to operate with qualities of objects without presuming a language on which these thoughts can be expressed. (See e.g. Gärdenfors 1990). This suggests that quality dimensions determine the conceptual space more or less absolutely.

Gärdenfors puts considerable effort into eliminating charges of relativism (2000, 81) and defends what we consider to be his variant of objectivism. According to him "our quality dimensions are what they are because they have been selected to fit the surrounding world" (Ibid, 83). This argument apparently addresses conditions that are determined by evolution and which may have existed before language emerged. Gärdenfors and Warglien (2007) assume that different individuals have

different mental spaces and thereby set out to solve the issue of shared semantics by the "meeting of minds" in terms of synchronized fixpoints. In Zenker's and Gärdenfors's discussion on conceptual change in scientific conceptual frameworks (2015) the idea of a given or fixed conceptual space is abandoned.

We interpret that Gärdenfors and his collaborators now see conceptual spaces as a means to study any conceptual structures, even abstract ones beyond the primordial level of cognition to which Gärdenfors (2000) appears to refer. As Gärdenfors and Williams (2001) discuss, a conceptual space is a flexible approach that can be modified in various ways. We follow this suggestion by introducing a *perspectivist* account of similarity, allowing the interactive exploration of alternative perspectives to the conceptual space.

13.2 Perspectivism

The recognition of the perspectival nature of cognition can be called *perspectivism*, following Giere's definition (2006). By means of an analogy of the spatial physical world, where objects appear in various ways depending on the perceiver's movements and points of view, even cognitive categories and concepts vary depending on the context or frame of reference. The approach has long historical roots, dating back at least to Protagoras and Heraclitus. Protagoras's maxim "*Man is the measure of all things*" sets the focus on the human agency of cognition, while Heraclitus' idea that "everything flows" introduces the essential dynamical aspect of perspectivism. It was Leibniz who first used the very term *perspectivism*, giving it a perceptual interpretation. According to his monadology, each individual, or "monad", perceives or mirrors the world from his own perspective. Perspectivism was later strongly associated with Friedrich Nietzsche, who In *Beyond Good and Evil* claimed that "there are no facts, only interpretations". Further, according to him, "*one always knows or perceives or thinks about something from a particular perspective – not just a spatial viewpoint, but a particular context of surrounding impressions, influences, and ideas, conceived of through one's language and social upbringing and, ultimately, determined by virtually everything about oneself, one's psychophysical make-up, and one's history*" (Solomon 1996, 195; See also Magnus and Higgins 1996).

Baghramian puts it that there can be more than one correct account of how things are in any given domain (2004, Chapter 10). If so, the issue is not which perspective is correct or true, but how to explore and mutually relate multiple perspectives. Consequently, there is no need to assume that the exploration of perspectives would at some point be satisfied, or to expect the convergence of perspectives to any final or 'true' form.

Similarly, in psychology Neisser and Jopling have suggested that categorization may well be based on similarity, but that similarity itself depends not only upon perceptual similarity but even involves "theory" (1997, 169). Neisser's perceptual cycle (1976) assumes a continuous systemic interaction between objects, their

perception and a cognitive schema, a kind of "theory". It is even empirically well-established that the judgment of similarity is all but deterministic (see e.g. Smith and Heise 1992, 242).

In philosophy of science, interpretations of observations are said to be theory-laden, that is, they depend on the theory adopted (e.g. Hanson 1958; Kuhn 1962; Feyerabend 1981), where 'theory' equals a particular perspective. Even the etymology of 'theory' supports this reading, with the Greek verb *theorein* referring to "to consider, speculate, and look at".[1] Another view on the multiplicity of perspectives is that of Pierre Duhem, the French scientist, who criticized the inductivism of Newton, stressing that "*An experiment in physics is not simply the observation of a phenomenon; it is, beside, the theoretical interpretation of this phenomenon*" (1962, 144). His framework, representing a kind of holism referred to as the *Duhem thesis*, expressed the following: "*An experiment in physics can never condemn an isolated hypothesis but only a whole theoretical group*" (ibid. 183.)

W.V.O. Quine later elaborated the argument, which thereby came to be known as the Duhem-Quine thesis (see Gillies 1993). He talked about "*the totality of our so-called knowledge or beliefs*" that is "*a man-made fabric which impinges on experience only along the edges*" (1980, 65). According to him, different theories, or as we may interpret them in the present context, conceptualizations, are underdetermined by experience and can be empirically equivalent. Thus, the same facts can support different, potentially inconsistent conceptualizations, each of which only partially matches the experienced reality.

Putnam's *pragmatic pluralism*, according to which the same things can be described in many different ways (see 2004), also borders perpectivism. In his linguistically oriented point of view, natural languages come with their own ontologies – entities that are talked about. He indicates that everyday language employs different kinds of discourses, subject to different standards and possessing different sorts of applications, with different logical and grammatical features – different language games (Ibid. 2004, 21–22, see also Rorty 1979).

A logical treatment of perspectivism was elaborated by Antti Hautamäki (1986), based on the concept of determinables, originally presented by Johnson in 1921 (1964). According to the Johnson, determinables are abstract names, adjectives, although grammatically they are substantival (color). Determinates or determinate values like different colors, in turn, produce logical divisions of the space of determinables. Thereby Hautamäki's study already implies the fundaments of a conceptual space.

[1] http://www.etymonline.com/index.php?term=theory

13.3 Ontospace Exploration Model

We originally introduced the perspectivist interpretation of Gärdenfors's conceptual spaces in Kaipainen and Hautamäki (2011), where we set the focus on interactive exploration of multiple perspectives during the process of conceptualization. This article focused on the variability of conceptualization (or categorization) as the function of perspectives to data taken interactively. We also related perspectives to short- and long-term contexts. Short-term contexts are constituted of narrative and situational factors and interpretative frames that are effective at the moment of observation. Long-term contexts may be as broad as natural conditions, evolution, or life-long learning. In the perspectivist spirit, the approach builds on the premise that there is no such thing as a concept without a perspective, but one is at least implicitly always present. This holds even for apparently absolute and neutral data, where there is an implicit perspective at least in the form of the choice and prioritization of determinables, applied metrics, or scalings.

Another key assumption we made is that perception and cognition, ultimately the brain, cannot effectively deal with unlimited dimensionality of the world since evolution has mainly adapted them to the constraints of the directly perceivable two- and three-dimensional aspects of the environment. Therefore, we generalize that the prerequisite of cognitive-perceptual sense making is to reduce the high (or infinite) dimensionality of the world, without feeling obliged to estimate the maximum dimensionality the cognitive-perceptual system can cope with. This is an empirical question that falls under the domain of psychology. The idea of dimensionality reduction was, of course, not unbeknownst to Gärdenfors in 2000 who put it: "going from the subconceptual to the conceptual level usually involves a reduction of the number of dimensions that are represented (221)". However, he chose not to elaborate this further as a part of his conceptual spaces model.

In order to be able to formalize the dimensionality reduction, we make a distinction between *ontospace* **A** and *representation space* **B**, constituting what we call the *ontospace model.* This allows us to study the dynamics of concept construction within a domain or discourse and, more particularly, to compare different conceptualizations concerning it. This formulation makes a distinction between the world or subject under discussion and the observer's perspectival interpretation of it. The ontospace represents the shared "world" constructed by joint observation, elaboration and research, but which instead of yielding to one shared conceptualization allows for a range of ways to describe the surrounding world. It can also be conceived of as a platform that allows the study of the dynamics of perspective exploration, interpersonal negotiation or deliberation between different perspectives, and the potential of higher-level knowledge beyond single perspectives emerging from the explorative activity.

Following Kaipainen and Hautamäki (2011), we start from the spatial metaphor of Gärdenfors and define *an ontospace* as a coordinate system describing a state space that specifies the dimensions with respect to which items of the topical domain vary. Let I be a set of Johnson (1921) determinables, corresponding to feature

dimensions in Gärdenfors's model. They may also be called attributes, features, properties or qualities in other contexts. To give an example of such a set, $I = \{$color, form, weight, length, ... $\}$. For now, we also assume that qualitative determinables can be transformed into quantitative variables, which is a standard procedure in measurement theory. Associated with each determinable i in I there is a set of determinate values D_i. Thus, an ontospace for a topic domain is an n-dimensional space $A = D_1 \times D_2 \times \ldots \times D_n$. Elements of A are n-tuples of the form $a = [a_1, a_2, \ldots, a_n]$, where a_i belongs to D_i. Each entity x of the topic domain can be represented as a state $s(x) = a_x$ in ontospace A, where $a_x = [a_{x1}, a_{x2}, \ldots, a_{xn}]$, of which the elements are also conceivable as the *ontocoordinates* of x. Note that s(x) determines the properties of x, assuming that properties are regions of the ontospace **A** and the state s(x) of x is a member of A. There is no need to assume A to be fixed. Rather, it can grow and shrink depending on the evolution of the discourse, culture, or scientific paradigm, whatever it is a model of.

Suppose that there is a distance measure m_i for all determinables i, expressing the degree of mutual similarity among elements in terms of set D_i of determinate values. Here m_i is a function from $A_i \times A_i$ to the set of non-negative real numbers R^+, where $m_i(a_i, b_i)$ is equal to the distance between values a_i and b_i in set A_i. Consequently, a larger distance means less similarity in a quality dimension. In terms of visualization, an ontospace is a multi-dimensional matrix that allows numerous agglomerative or divisive hierarchical clustering algorithms to be applied, such as multidimensional scaling MDS (e.g.Kruskal and Wish 1978), Kohonen's self-organizing map SOM (Kohonen 1982), principal component analysis PCA, or *Eigentaste* (Goldberg et al. 2001), insofar as they allow the representation of data elements of A in a representation space B of lower dimensionality while maintaining similarity relations in A. The only condition is that the applied algorithm needs to allow weighting or prioritization by means of the additional element perspective P, determining the dimensions with respect to which similarity relations are to be prioritized. Thus, P is a means of expressing relativity of the similarity relations in A.

A *perspective* to ontospace A is defined as an array $P = [p_1, p_2, \ldots, p_n]$ of weights, for all determinables i. Following Kaipainen et al. (2008) we assume that a perspective applies to a *selection of determinables* as in the treatment of Hautamäki (1986), but in this case allowing all real numbered values ranging within interval [0,1]. The weight p_i expresses the *interest* or *attention* of an observer towards the ontocoordinate i. A central notion of our approach is the transformation R_P, called *reduction function*, from the high-dimensional ontospace A to the lower-dimensional representational space B. The perspective P has the role of constraining R_P. As a prerequisite for this transformation, we generally assume a distance measure M in B, corresponding to similarity from the viewer's viewpoint. It can defined in several alternative ways, including Euclidean distance, street block distance, and as a more general formulation, the Minkowski metric.

A reduction function R_P from A to B respects the perspective P and distance measures in the following way:

(a) If $p_i = 1$, then the distance m_i contributes fully to the distance measure M

(b) If $p_i = 0$ then the distance m_i is ignored by M.
(c) Intermediate values $0 < p_i < 1$ refer to partial contributions to the distance measure M.

By means of the function R_P, objects of the domain can be categorized in a manner that reflects the adopted perspective. The exact character of the reduction function needs not to be fixed, the only constraint being that R_P is sensitive to values p_i according to conditions (a), (b), and (c). The elements p_i of P function as *weights* a viewer gives to ontocoordinates. Zero means a total ignorance of the dimension in question.

As previously discussed, our assumption is that dimensionality-reduced mapping R_P facilitates the cognitive manageability of A. Here the ontological space A is interpreted in terms of the representational space B, in particular, similarity relations in A are observed by means of similarity relations in B. Mathematically, a reduction function R_P induces similarity relations in A based on similarity relations in B: if $R_P(a)$ and $R_P(b)$ are similar in B then a and b are considered to be similar. This way similarity in A is relativized to perspectives P. Thus perspectives P regulate the spatial clustering and its interpretation as a conceptual space. In our perspectival approach, the representational space B implies a conceptual organization of the items associated with the A. Depending on the adopted theoretical approach, the organization can be conceived of as in terms of *prototypes*, *categories*, or *tessellations* (see Gärdenfors 2004), or even *mereological relations*, in every case assuming they are perspective-relative.

Thus in our approach, the number of dimensions cannot only grow towards infinite but can also diminish dynamically over time. One can interpret that concepts (and conceptualizations) are constructed on the fly, as seen from the currently relevant perspective that reflects the particular priorities, interests, and contextual conditions relevant in the particular point of time for the particular cognizer or community.

Thus, this approach implies two ways of expressing relativity. The construction of ontospace is dependent on its cultural context, reflecting what is possible to know in the present state of the knowledge. On the other hand, the construction of representational spaces is relative to the particular viewers' conditions. The major impact of the differentiation of ontospace and representational space is that it allows interactively dynamic explorations and comparisons of conceptualizations of the same data.

13.4 A Case Study

In terms of a quasi-Linnaean example, let's assume a corpus of fauna consisting of *dog*, *pig*, *human*, *gorilla*, *elephant*, *snake* and *crow*, each occupying an ontospace determined by coordinates corresponding to the following ontodimension (property dimensions): number of *legs*, thickness of hairy *skincover*, *weight*, *intelligence*, and

Table 13.1 The data Creatures with columns indicating the property-describing coordinates of each item within the ontospace

	Legs	Skincover	Weight	Intelligence	Speed
Dog	1.000	0.900	0.322	0.556	0.917
Pig	1.000	0.500	0.407	0.444	0.167
Human	0.500	0.100	0.525	1.000	0.333
Gorilla	0.500	1.000	0.576	0.889	0.500
Snake	0.000	0.000	0.068	0.000	0.833
Elephant	1.000	0.100	1.000	0.889	0.000
Crow	0.500	0.800	0.000	0.444	1.000

speed (Table 13.1). In the example data, all dimensions are scaled to range between 0 and 1. Thus, 1 corresponds to the maximum number of legs (four), and 0 to the minimum (no legs).

Unlike Linnaeus' fixed taxonomies of fauna, the perspectivist approach allows the corpus to be conceptualized not only in one way but also in a number of different ways, depending on the perspective chosen by the observer. Let's consider two examples of alternative perspectives and corresponding conceptualizations. Obviously, concept names are not directly derived from the model itself, but must rely on some linguistic or cultural convention outside of the current topic. In this particular example, concepts are identified with the most typical member of the concept, i.e. the mean.

From the perspective depicted in Fig. 13.1, the domain Creatures is conceptualized in the following manner. Here, humans, pigs and gorillas together constitute a convex cluster corresponding to the concept of the '*human-like*', characterized by high intelligence, low weight, and average speed.

Zoomed out from the same perspective (Fig. 13.2), they, in turn, belong to the broader concept of the '*gorilla-like*', the intelligent ones being distinguished from the non-intelligent *'snake-like'*. Within the '*gorilla-like*', the '*human-like*' are separated from those labeled '*dog-like*', '*crow-like*' and the '*elephant-like*', with respect to the particular properties weighted by the perspective.

From yet another perspective (Fig. 13.3), humans would be associated with the concept '*gorilla-like*'.

In sum, the approach allows the epistemic exploration of a conceptual space from multiple perspectives to the same data in the perspectivist sense, giving rise to corresponding hierarchically embedded conceptualizations that satisfy the convexity condition stipulated by Gärdenfors.

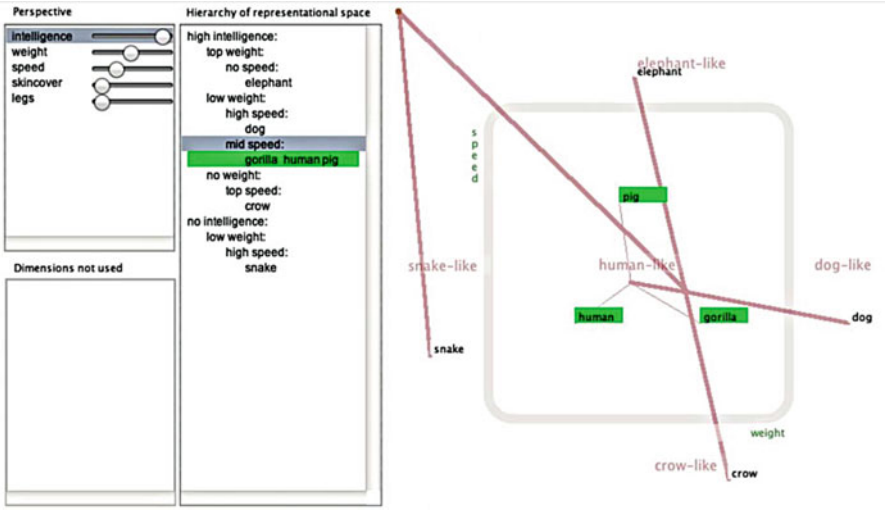

Fig. 13.1 A tree structure depicting the hierarchical conceptualization of the *Creatures* from a perspective of weights intelligence (1.0), skincover (0.5), weight (0.33), legs (0), and speed (0), as controlled by the slider positions in the *left panel*. The corresponding hierarchical structure of the representational space is schematized textually as an embedded list in the *middle panel*

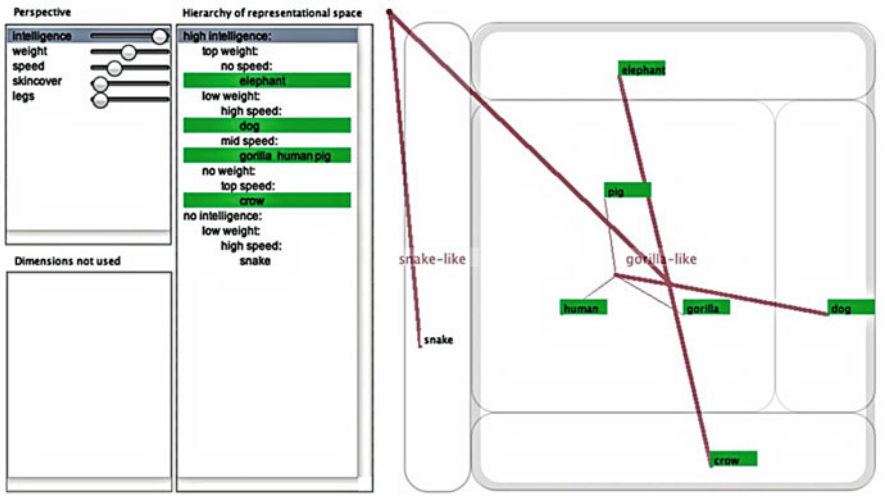

Fig. 13.2 A broader conceptualization zoomed out from the same perspective as in Fig. 13.1

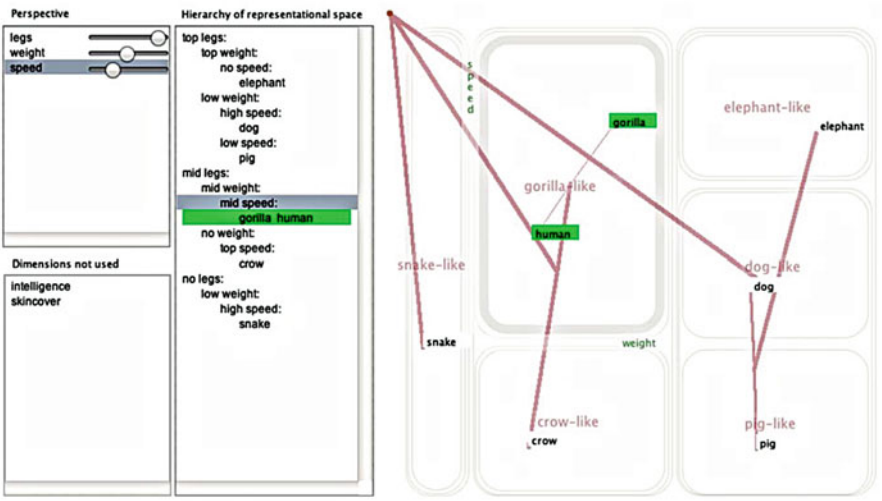

Fig. 13.3 Perspective with the number of legs weighted (1.0), weight (0.5) and speed (0.33), associating humans with the concept of '*gorilla-like*'

13.5 Discussion

Similarity is not given, it is "similarity-for-us", as Popper once said (1953, 45). The suggested model aims to explain the perception of similarity from different perspectives. The association of similarity with concepts, as assumed by the conceptual spaces paradigm among others, has allowed us to talk about multi-perspective exploration of concepts. However, it is not *concept* as some kind of a static construct but rather the capability of explorative *conceptualization* that is in focus. The model suggests that conceptualization of abstract entities is reminiscent of observing artifacts in a physical space, where the appearance of the artifacts depends on the observer's distance and angle to it. There is no single two-dimensional image of a chair that would suffice to exhaust the nature of such an object, but the full understanding of a chair requires multiple perspectives of it. As in the case of exploring physical objects, there is no need to assume abstract conceptualization to be static.

It is important to emphasize that an ontospace is not "given" by the world, neither is it a perspective-independent representation of the "world". It can be seen as a culturally constructed 'archive' of all dimensions possible in the context of discussion, referring to Foucault's term (1972). Essentially, the conceptualization of the "world", manifested in the representational space, is relative to the perspective adopted.

In this context the Gärdenforsian conceptual space can be conceived of being underspecified,[2] referring to both an ontospace and to an implicit perspective, under which every dimension is equally given the value 1, non-sensitive to all choices, prioritizations, scalings that have taken place during the accumulation and pre-processing of data. Our model merely makes explicit that which is implicit in conceptual spaces.

The ontospace exploration approach has already proved its feasibility in several fields of application. In Kaipainen and Hautamäki (2011) we propose an interactive application aimed to facilitate the exploration of multiple perspectives in the field of knowledge organization. It allows the user to explore a topical domain (modeled as an ontospace) from multiple perspectives in order to construct alternative conceptualizations, thereby implying a kind of multi-perspective medium. One of the obvious application areas of ontospaces is to model the accumulation of narrative contexts during unfolding stories in narrative arts. A narrative ontospace is continuously accumulated as the story unfolds. A narrative perspective assigns weights or priorities to the narrative dimensions that are brought into focus at a particular time. This concept has been applied in the field of interactive cinema (Tikka et al. 2006; Tikka 2008; Pugliese et al. 2014). Elsewhere, a computational model of ecological learning dynamics has been proposed, based on the foundation of an explorable ontospace (Normak et al. 2012). Yet, in another direction, the ontospace exploration model has been applied as an interactive approach to clustering techniques in data analysis and visualization within the *mixed methods* paradigm (Niglas and Kaipainen 2008).

With regard to the present context, Zenker and Gärdenfors suggest that *change in importance* [of dimensions] and *addition and deletion of dimensions* characterize shifts in conceptual frameworks of science (Chap. 14, this volume). Our model can accommodate such changes, interpreted as shifts of perspective to the implied ontospace of such frameworks (cf. Hautamäki 1986, Chapter V).

Apart from epistemic exploration of ontospaces in the service of an individual cognition, perspective-relative conceptualization can be seen as a model of how an individual can simulate another person's conceptualization of a domain, that is, to simulate another person's point of view. One possibility is to make perspectives as interchangeable media items, allowing the concretization of what is referred to as 'my perspective' or 'their point of view'. Taken further, multi-perspective conceptualization may serve as a model of *deliberation* in political or ethical domains, complying with the Deweyan tradition (see e.g. Caspary 2000), or *dialogization* of texts in the sense of Bakhtin (Holquist 1981). One may also foresee the advantages of being able to point out the perspective-dependent nature of complex and ambiguous ethical, political or philosophical domains and guide the receivers to explore the alternative views on their own instead of an authoritative perspective. Quite obviously, this bears particular potential for educational purposes.

[2]Peter Gärdenfors, personal communication (2012).

The model not only accommodates perspective-dependence and explorativeness with Gärdenfors's conceptual spaces model, but in the bigger picture it is compatible with all similarity-based conceptualization approaches, like those presented under the label "embodied realism" by Lakoff and Johnson (1999, Chapter 6). It also turns some of the criticism raised against similarity as the basis of conceptualization to a favor (Gärdenfors 2000, 109). The relative nature of similarity does not undermine its significance for cognition. On the contrary, it makes it compelling to analyze the factors with respect to which similarity is relative. In addition to the remarks discussed earlier, Murphy and Medin (1985, 291) have pointed out that "*at its best, similarity only provides a language for talking about conceptual coherence*". However, is not that what concepts are for? The language they propose would alone suffice as a cause for celebration, since it is exactly the communicability and shareability offered by perspective-dependent conceptualization that allows advanced intersubjectivity and consensus.

Nevertheless, we leave it to future discussions to determine to which degree the introduced model can cover the scope of meanings that have been and can be associated with the idea of concepts. Gauker's claim that similarity spaces cannot model concepts (2007) represents a conflicting definition of concepts and must therefore be left to further discussions.

Obviously, the presented model must be further elaborated to provide a more complete account of the dynamics, patterns, sequences or strategies that lead to higher understanding over time. The way to such a level goes via concrete cases and system dynamics associated with them.

Acknowledgements We thank Södertörn University, The Foundation for Baltic and East European Studies, and Jyväskylä University for making this work possible, Peter Gärdenfors, Frank Zenker, and the anonymous referees for a number of good hints and suggestions.

References

Baghramian, M. (2004). *Relativism*. London/New York: Routledge.

Caspary, W. R. (2000). *Dewey on democracy*. Ithaca: Cornell University Press.

Duhem, P. (1962). *The aim and structure of physical theory*. English translation by Philip P. Wiener of 2nd the French edition of 1914. Original French edition 1904–5. Atheneum.

Feyerabend, P. (1981). *Problems of empiricism: Philosophical papers, volume 2*. Cambridge: Cambridge University Press.

Foucault, M. (1972). *The archeology of knowledge*. New York: Pantheon Books.

Gärdenfors, P. (1990). Induction, conceptual spaces and AI. *Philosophy of Science, 57*, 78–95.

Gärdenfors, P. (1992). A geometric model of concept formation. *Information Modelling and Knowledge Bases, 3*, 1–16.

Gärdenfors, P. (2000). *Conceptual spaces: On the geometry of thought*. Cambridge, MA: The MIT Press.

Gärdenfors, P. (2004). Conceptual spaces as a framework for knowledge representation. Imprint Academic. *Mind and Matter, 2*(2), 9–27.

Gärdenfors, P., & Warglien, M. (2007). *Semantics, conceptual spaces, and the meeting of minds*. LUCS Cognitive Science Centre, University of Lund, Lund, Sweden.

Gärdenfors, P., & Williams, M.-A. (2001). *Reasoning about categories in conceptual spaces*. Proceedings international joint conference on artificial intelligence. Seattle, WA, USA.

Gärdenfors, P., & Zenker, F. (2015). Communication, rationality, and conceptual changes in scientific theories. In F. Zenker & P. Gärdenfors (Eds.), *Applications of conceptual spaces: The case for geometric knowledge representation* (Synthese Library). Dordrecht: Springer.

Gauker, C. (2007). A critique of the similarity space theory of concepts. *Mind & Language, 22*(4), 317–345.

Giere, R. N. (2006). *Scientific perspectivism*. Chicago/London: University of Chicago Press.

Gillies, D. (1993). *Philosophy of science in the twentieth century*. Cambridge, MA: Blackwell.

Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2001). Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval, 4*(2), 1386–4564.

Goodman, N. (1978). *Ways of worldmaking* (Vol. 51). Indianapolis: Hackett Publishing.

Hanson, N. R. (1958). *Patterns of discovery*. Cambridge: Cambridge University Press.

Hautamäki, A. (1986). *Points of view and their logical analysis* (Acta Philosophica Fennica 41). Helsinki: Societas Philosophica Fennica.

Holquist, M. (Ed.). (1981). *The dialogic imagination. Four essays of M.M. Bakthin*. Austin: University of Texas Press.

Johannesson, M. (2002). *Geometric models of similarity*. Lund: Lund University Cognitive Studies.

Johnson, WE. (1921). Logic, 3 vols. Cambridge, UK: Cambridge.

Kaipainen, M., & Hautamäki, A. (2011). Epistemic pluralism and multi-perspective knowledge organization, explorative conceptualization of topical content domains. *Knowledge Organization, 38*(6), 503–514.

Kaipainen, M., Normak, P., Niglas, K., Kippar, J., & Laanpere, M. (2008). Soft ontologies, spatial representations and multiperspective explorability. *Expert Systems, 25*(5), 474–483.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43*, 59–69.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Thousand Oaks: Sage.

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh, the embodied mind and its challenge to western thought*. New York: Basic Books.

Magnus, B., & Higgins, K. M. (Eds.). (1996). *The Cambridge companion to Nietzsche*. Cambridge: Cambridge University Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *The Psychological Review, 92*, 289–316.

Neisser, U. (1976). *Cognition and reality. Principles and implications of cognitive psychology*. San Francisco: W. H. Freeman and Company.

Neisser, U., & Jopling, D. A. (1997). *The conceptual self in context: Culture, experience, self-understanding*. Cambridge: Cambridge University Press.

Niglas, K., & Kaipainen, M. (2008). Multi-perspective exploration as a tool for mixed methods research. In M. Bergman (Ed.), *Advances in mixed methods research: Theories and applications* (pp. 172–188). Los Angeles/London: Sage.

Normak, P., Pata, K., & Kaipainen, M. (2012). An ecological approach to learning dynamics. *Educational Technology & Society, 15*(3), Special issue on Learning and Knowledge Analytics, 262–274.

Popper, K. (1953). *Conjectures and refutations, the growth of scientific knowledge*. London: Routledge and Kegan Paul.

Pugliese, R., Tikka, P., & Kaipainen, M. (2014, November 1). *Navigating story ontospace: Perspective-relative drive and combinatory montage of cinematic content*. Studies on art and architecture. Special issue on expanding practices in audiovisual narrative.

Putnam, H. (2004). *Ethics without ontology*. Cambridge, MA: Harvard University Press.

Quine, W. V. O. (1980). Two dogmas of empiricism. In H. Morick (Ed.), *Challenges to empiricism* (pp. 46–70). London: Methuen.

Rorty, R. (1979). *Philosophy and the mirror of nature*. Princeton: Princeton University Press.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology, 4*, 328–350.

Rosch, E. H. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*, 192–232.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, New Series, 237*(4820), 1317–1323.

Smith, L. B., & Heise, D. (1992). Perceptual similarity and conceptual structure. In B. Burns (Ed), *Percepts, concepts and categories* (pp. 233–233). Elsevier Science Publishers.

Solomon, R. C. (1996). Nietzsche ad hominem: Perspectivism, personality and ressentiment. In B. Magnus & K. M. Hggins (Eds.), *The Cambridge companion to Nietzsche* (pp. 180–222). Cambridge/New York: Cambridge University Press.

Tikka, P. (2008). *Enactive Cinema. Simulatorium Eisensteinense*. Publication series of the University of Art and Design Helsinki.

Tikka, P., Vuori, R., & Kaipainen, M. (2006). Narrative logic of enactive cinema: Obsession. *Digital Creativity, 17*(4), 205–212.

Chapter 14
Communication, Rationality, and Conceptual Changes in Scientific Theories

Frank Zenker and Peter Gärdenfors

Abstract This article outlines how conceptual spaces theory applies to modeling changes of scientific frameworks when these are treated as spatial structures rather than as linguistic entities. The theory is briefly introduced and five types of changes are presented. It is then contrasted with Michael Friedman's neo-Kantian account that seeks to render Kuhn's "paradigm shift" as a communicatively rational historical event of conceptual development in the sciences. Like Friedman, we refer to the transition from Newtonian to relativistic mechanics as an example of "deep conceptual change." But we take the communicative rationality of radical conceptual change to be available prior to the philosophical meta-paradigms that Friedman deems indispensable for this purpose.

14.1 Introduction

Thomas Kuhn (1970) famously argued that successive paradigms in the sciences are incommensurable. It is therefore a challenge for philosophers of science to explain how adherents of successive paradigms can nevertheless engage in rational communication during periods of "deep conceptual change." The transition from Newtonian to relativistic mechanics is widely accepted as *the* paradigmatic example. Michael Friedman's (2001, 2002a, b, 2008, 2010) strategy in addressing this challenge is to divide the scientific discourse into three levels. In order to account for rational communication about competing theory frameworks (paradigms), he argues that meta-philosophical arguments are required.

Common to Kuhn's and Friedman's positions is the assumption that scientific theories are *linguistic entities* and that successful communication about them implies a realist view of semantics whenever theories predict natural phenomena. As a theory's terms—for instance, *mass*—then apply by *reference* to things—for instance, planets—communicative impediments would appear to arise because of the differences in how referents are construed across paradigms. Our aim in

F. Zenker (✉) • P. Gärdenfors
Department of Philosophy and Cognitive Science, Lund University, Box 192, 221 00 Lund, Sweden
e-mail: frank.zenker@fil.lu.se; peter.gardenfors@lucs.lu.se

F. Zenker, P. Gärdenfors (eds.), *Applications of Conceptual Spaces*,
Synthese Library 359, DOI 10.1007/978-3-319-15021-5_14

this article is to present an alternative way of meeting Kuhn's challenge. Our strategy is to see a theory as a constraint on a spatial structure (the framework of a theory). We model such frameworks in terms of conceptual spaces (Gärdenfors 2000; Gärdenfors and Zenker 2014). By analyzing different kinds of changes within such spatial structures, one can understand Kuhn's challenge in a way that eschews incommensurability and that does not require meta-philosophical discussions.

The communicative impediments that arise in deep conceptual change may thus be alleviated by focusing on the structural representation of concepts in conceptual spaces instead of their possible referents. This shifts focus away from referents and towards the structure of concepts. The theory of conceptual spaces, then, is employed as a particular example of a more general answer to Kuhn's challenge. We also believe that our perspective lies closer to what scientists actually use in their thinking and communication.

The paper proceeds as follows: we present how conceptual frameworks can be understood as spatial entities (Sect. 14.2), how the dynamics of such frameworks may be analyzed in the diachronic case (Sect. 14.3), and how to address communicative challenges that are caused by meaning change (Sect. 14.4). We then contrast our approach with Michael Friedman's neo-Kantianism (Sect. 14.5), and argue that an explanation of rational communication between proponents of successive paradigms need not rely on meta-philosophical paradigms (Sect. 14.6).

14.2 Conceptual Frameworks as Spatial Entities

The theory of conceptual spaces builds on a cognitivist view of semantics. It contrasts with both extensional and intensional realistic semantics that include the referent of a linguistic expression as a meaning constituent. On our approach, conceptual knowledge is seen as mental representations, modeled in conceptual spaces. We do not view such representations as parts of a symbolic system with a syntactic or logical structure. Instead, we treat them as spatial structures that can be analyzed into their constitutive dimensions and properties, representing the semantic knowledge of an agent (Gärdenfors 2000, 2014).

An empirical theory always presupposes a specific conceptual framework that provides the magnitudes, or dimensions, on which the formulation of this theory depends. These magnitudes can be modeled as collections of dimensions with their inter-relations, that is, as conceptual spaces (Gärdenfors 2000; Gärdenfors and Zenker 2011, 2013; Zenker and Gärdenfors 2014; Zenker 2014). Put schematically, an empirical theory, T, depends on a conceptual framework, F, that is modeled as a conceptual space, S.

Apart from concepts arising in sensory perception, e.g., color or sound, or in basic scientific measurements, magnitudes include those introduced by science, for instance, *mass*, *force*, and *energy*. The topological or metrical structure of such magnitudes is tightly connected to the methods by which they can be measured, and to their relations to other concepts within a scientific theory. *Newtonian mass*,

for instance, can be modeled as a dimension that is isomorphic to the non-negative part of the real number line, and Newtonian space as a three dimensional integral (vector) space with Euclidian metric. (We return to this example below.)

The primary role of the dimensions is to represent various "qualities" of objects in different *domains*. The notion of a domain can be given a more precise meaning by using the notions of *separable* and *integral* dimensions. Dimensions are said to be integral if, to describe an object fully, one cannot assign it a value on one dimension without giving it a value on the other. For example, an object cannot be given a hue without also giving it a brightness value. Or the pitch of a sound always goes along with its loudness. Dimensions that are not integral are said to be *separable*, as for example the size and hue dimensions. Within the context of scientific theories, the distinction should rather be defined in terms of *measurement procedures*. If two dimensions (or sets of dimensions) can be measured by independent methods, then they are separable, otherwise they are integral.

It is therefore part of the meaning of "integral dimensions" that they share a metric, which separable dimensions do not. A theory *domain* can now be defined as the set of integral dimensions that are separable from all other dimensions in a theory, for instance Newton's domain of absolute space.[1]

The choice of domains need not be uniquely determined, for the organization of dimensions into domains depends on which dimensions are considered basic for a theory. In the literature one finds organization principles that are theoretically motivated, e.g., via transformation classes, or motivated by practical considerations, e.g., via measurement procedures. The set of domains thus depends on the choice of basic dimensions (notably the basic ones given by the International System of Units—SI-units), and the choice of useful derived dimensions. Both ontological and epistemological considerations have historically informed which dimensions are considered basic and which are derived, but our notion of what constitutes a domain is an instrumentalist one. This notion is independent of ontological positions, and while it is compatible with calling certain dimensions basic or fundamental that are found to be epistemologically prior, no such distinction is required.[2]

[1] More precisely, domain C is separable from D in a theory, if and only if the invariance transformations of the dimensions in C do not involve any dimension from D; and the dimensions of a domain C are integral, if and only if their invariance class does not involve any other dimension (Gärdenfors and Zenker 2013).

[2] Albert Einstein, for instance, appears to have been well aware of the Kantian tradition, yet opposed to singling out some concepts as privileged or epistemologically prior:

> "One must not … speak of the 'ideality of space.' 'Ideality' pertains to all concepts, those referring to space and time no less and no more than all others. Only a complete scientific conceptual system comes to be univocally coordinated with sensory experience. On my view, Kant has influenced the development of our thinking in an unfavorable way, in that he has ascribed a special status to spatio-temporal concepts and their relations in contrast to other concepts." (Einstein 1924, 1690f.; cited after Howard (2010, 347)

Contrasting with our instrumentalist notion, Kuhn's incommensurability thesis derives from a view on semantics that requires all concepts to have referents. His paradigmatic example in support of the thesis is that changes in meaning of the scientific variables (dimensions) go along with a change of referent. He writes:

> "The variables and parameters that in [the statements which Kuhn refers to as] the Einsteinian E_i's represented spatial position, time, mass, etc. still occur in the [Newtonian] N_i's and they still represent Einsteinian space, time and, mass. But the *physical referents* of these Einsteinian concepts are by no means identical with those of the Newtonian concepts that bear the same name. (Newtonian mass is conserved; Einsteinian is convertible with energy. Only at low relative velocities may the two be measured in the same way, and even then they must not be conceived to be the same.)" (Kuhn 1970, 101f.; italics added)

And it is vis-à-vis such a semantic realism that Kuhn can claim that the predecessor and the successor theory in a paradigm shift are non-intertranslatable.

Since our position is basically instrumentalist, we do not count a theoretical term such as *mass* as one that has a direct reference in an external world. Consequently, we can analyze, and then compare, different mass concepts solely in virtue of their distinct dimensional reconstitutions, and their relations to other concepts.[3] *Newtonian mass*, for instance, is separable from everything else in Newtonian theory; it constitutes a separate dimension because an object's mass remains constant, and so is independent of variations in other magnitudes such as the object's position or velocity. *Relativistic mass*, in contrast, is integral with *energy* in special relativity: the mass of a moving object increases as its kinetic energy increases, given the object has non-zero rest mass. The term *mass* thus takes on different meanings in these two theories. Both meanings can nevertheless be compared—something Kuhn was well aware of (d'Agostino 2013). But, as we will argue, it is misleading to present the meaning difference as giving rise to communicative difficulties of the kind that Kuhn took to support his claim according to which proponents of successive paradigms *cannot* engage in rational discourse—and so they *must* resort to less rational forms of interaction (see below).

Within a cognitive theory of meaning, as we have stated, the meaning of *mass* is analyzed non-realistically through a dimensional (spatial) representation. This view of the semantics of a scientific concept can be extended so that the dynamics of a scientific conceptual framework may be treated in like manner. We do this by analytically separating the types of changes that modify a predecessor framework into a successor framework. The term *framework* we here understand as the conceptual space that a scientific theory builds on.

Applying this typology yields a systematic analogue to the actual historical dynamics of theory frameworks insofar as a predecessor conceptual space can be

[3]Conversely, allegedly different concepts are in fact not distinct when sharing the same meaning in virtue of their identical dimensional forms and identical relations to other concepts. Thus, conceptual spaces may be particularly helpful to structural realists (Ladyman 2014) who seek to account for various aspects of continuity among theoretical structures—which may hence be *real*—as empirical theories formulated against the background of such frameworks undergo historical change.

modified into a successor space. As we argue below, it becomes plausible that, by understanding scientific concepts as spatial entities, scientists may successfully and rationally communicate about different concepts, their non-referentially conceived meanings, and thus about the dynamics of empirical theories in which these concepts are used. The threat of incommensurability is thereby eschewed. Before considering the arguments for this, we first turn to the types of changes that modify a conceptual framework.

14.3 The Dynamics of Conceptual Frameworks for Theories

In earlier work, we have provided a range of historical examples of five types of changes to a scientific conceptual framework (Gärdenfors and Zenker 2011, 2013). Here we restrict ourselves to a brief overview. In order of increasing severity of change, the following are distinguished: addition and deletion of special laws; change of metric; change in importance; change in separability; addition and deletion of dimensions.

Firstly, in our model, the special laws of a theory provide constraints on the distribution of points over a conceptual space. Newton's second axiom, for instance, restricts points to the hyper-surface described by $F = ma$ in the conceptual space consisting of the domains *time*, *mass*, (physical) *space*, and *force* (Gärdenfors 2000; Gärdenfors and Zenker 2013). The law of gravitation further restricts these points to regions where $F = GMm/r^2$ holds (Zenker 2009, 41f.). The addition or deletion of axioms and special laws is the mildest type of change, as it leaves the conceptual framework intact.

Secondly, each domain is endowed with a scale or metric that determines distances. For instance, day and nighttime are standardly modeled by one circular dimension with 24 equally-sized intervals called *hours*. Before the invention of mechanical clocks, however, the two points on this dimension that separate 12 h of nighttime from 12 h of daytime were commonly coordinated *locally* to sunrise and sunset. As these points shift, again locally, over the course of one year, their distance changes.[4] This variable temporal metric has meanwhile given way to one of constant clock intervals. The same occurred in the case of space—yet in the inverse direction, namely from constant to variable—where the Euclidian metric assumed by Newton first gave way to a Minkowskian (special relativity) and then to a Schwarzschild metric (general relativity), which may both be viewed to generalize the Euclidian. It is easy to see that a change of metric leads to a change in the symbolic expressions used to calculate with these distances.

[4]Consequently, the expression "days are longer in summer than in winter" was literally true although nights and days were always 12 h long. But the length of an hour varied from day to day, ever more markedly so over the contrasting seasons at locations further away from the equator.

Thirdly, the extant dimensions (or domains) provided by a scientific framework may change in importance. Energy, for instance, was initially of little importance to Newtonian mechanics (Gärdenfors and Zenker 2013). In contrast, forces became ever more important in the development of nineteenth century fluid dynamics, among others through a separation into short and long range forces (Petersen and Zenker 2014).

Fourthly, more severe yet are changes in the separability of dimensions, for instance the integration of mass with energy in relativistic physics, discussed above, or the formation of integral 4D space-time.

Fifthly, the most radical type of change is the addition to or the deletion of a dimension from a scientific framework as in the addition of the dimension *charge* in electrodynamics, or the abandonment of a dimension known as *caloric* in early accounts of thermo-dynamical phenomena.

The first of the five types addresses *intra*-framework changes; the other four types describe *inter*-framework changes. Categorizing changes of theories into these five types provides a finer grain than Kuhn's distinction between normal and revolutionary change. This gives us a richer toolbox to analyze the changes of a theory or its associated framework over time (for a case study see Gärdenfors and Zenker 2013). What is commonly referred to as *scientific revolution*, we categorize primarily under the last type: the replacement of dimensions.

14.4 Meaning Change and Communicative Challenges

This typology of scientific changes in hand, one can now address cases where a predecessor conceptual framework is replaced by a successor framework, such that the latter framework retains terms already used by the former. In short, one can address the kind of meaning change said to be typical of a "scientific revolution." In contrast to proponents of realistic semantics, our typology of changes allows to avoid the commitment that such frameworks (and the theories expressed relative to them) are non-intertranslatable because their shared terms diverge referentially. Since external referents are not central to our cognitivist account, the communicative challenges that proponents of pre- and post-revolutionary theories might face do consequently *not* arise from referential divergence, and so can have little to do with issues of realism. Such divergence, therefore, does not pose a threat to communicative rationality. With respect to Kuhn's main example, even if the meaning of *mass* is different in Newtonian mechanics and Einstein's relativity theory, it does not follow that the two meanings cannot be compared, since the geometric and topological properties of the Newtonian and Einsteinian frameworks still have a considerable overlap. Nor does it follow, provided such overlap but without referential stability, that proponents of successive paradigms must engage in broadly non-rational forms of communication.

Put differently, even if communicative difficulties are expectable in periods of deep conceptual change,[5] these impediments may be alleviated by focusing on the structural representation of concepts in conceptual spaces instead of their possible referents. Historically, it seems that scientists' communicative difficulties do not pertain to the role of the central concepts in scientific reasoning, but rather to unrelated factors such as their ontological commitments or research traditions.

We believe that an understanding of conceptual frameworks as dimensionally structured spaces also suffices to explain cases of successful and rational communication between scientists. Our view nevertheless remains at odds with Michael Friedman's account who postulates various explanatory levels to render cases of severe conceptual change as communicatively rational.

14.5 Michael Friedman's Neo-Kantianism

Having outlined conceptual spaces as a non-linguistic model of conceptual knowledge, we now turn to a presentation of the central tenets of Michael Friedman's influential Neo-Kantian account. It presents a modern version of Kantian epistemology that intends to render scientific revolutions as "communicatively rational" instances of scientific development. Friedman borrows this term from Jürgen Habermas's (1984, 1987) theory of communicative action where it broadly signifies the "non-coercively uniting, consensus creating power of argumentative speech" (Habermas, cited after Friedman 2001, 54).

Communicative rationality here contrasts with instrumental rationality. The latter governs the strategic choice among extant means, options, or tools, in order to satisfy, or optimize on, goals, purposes, or criteria, broadly conceived. Applied to the scientific case, for instance, instrumental rationality pertains to choosing among a range of available candidates that theory which is simpler, more empirically adequate, more fruitful, has greater scope, etc. A theory thus selected standardly constitutes an instrumentally rational choice with respect to the relevant goal or criterion, *only if* it is not dominated by another theory (maximization of choice utility).

Communicative rationality, in contrast, pertains—or, as the case may be, fails to pertain—not to the selection among *extant* means or tools, but to the communicative process that makes "the serious consideration of [a theory that only arises with] the *new* paradigm a rational and responsible option" (Friedman 2002a, 190, italics added; see his 2010, 714). So communicative rationality refers to the rational

[5] Rudolph Carnap, Kuhn's editor for *Structure* (see Reisch 1991), for instance, states:

> "I have found that most scientists and philosophers are willing to discuss a new assertion, if it is formulated in the customary conceptual framework; but it seems very difficult to most of them even to consider and discuss new concepts." (Schilpp 1963, 77)

quality of the interaction—itself antecedent to theory choice—that renders the new paradigm (e.g., relativistic mechanics) acceptable as a strategically rational option, and more particularly as an *additional option* for those scientists (e.g., the "Newtonians") who had formerly adhered strictly to the old paradigm only.

An immediate consequence of Friedman's view is that coercive or otherwise broadly non-rational forms of communication need not be postulated. For whenever such a process can in fact be communicatively rational in the sense of instantiating a form of non-coercive rational persuasion, then it makes little sense to suggest, as Kuhn (1970, ch. IX) did,[6] that the conversion of "old paradigmers" be a matter of *merely assent-directed* persuasion, thereby leveling the distinction between rational consensus and mere consensus.[7]

This being the purpose for which Friedman employs his Neo-Kantian account, we now turn to his notion of the relativized a priori (Sect. 14.5.1) and the role of philosophical meta-paradigms in deep conceptual change (Sect. 14.5.2).

14.5.1 The Relativized a Priori

Immanuel Kant had declared Newtonian time and space as a priori forms of human thinking, particularly an absolute space with Euclidian metric, and Newton's notion of absolute time. Because these forms make measurements in the empirical world possible, they are treated as *constitutive* of empirical knowledge. The a priori forms can also be expressed linguistically as principles. They then become, for instance, "the Newtonian Laws of Motion in the context of the *Principia*, the light principle and the principle of relativity in the context of [Einsteinian] special relativity, [or] the light principle and the principle of equivalence in the context of general relativity ..." (Friedman 2010, 713).[8] The a priori forms and their corresponding principles are taken to enable the kind of knowledge that empirical theories express

[6] Addressing the "genetic aspect of the parallel between political and scientific development" that "should no longer be open to doubt," Kuhn writes:

> "As in political revolutions, so in paradigm choice—there is no standard higher than the assent of the relevant community. To discover how scientific revolutions are effected, we shall therefore have to examine not only the impact of nature and of logic, but also the techniques of persuasive argumentation effective within the quite special groups that constitute the community of scientists." (Kuhn 1970, 94)

[7] Hence, securing the communicative rationality of paradigm shifts is also an attempt at criticizing the relativistic inclinations of the strong program in the sociology of scientific knowledge ("SSK program") (Rehg 2009, 14–80; see Friedman 1998).

[8] Newton's laws of motions standardly comprise *inertia*, $F = ma$ and, on many accounts, also *actio* = *reactio*. For general and special relativity, the *light principle* dictates that the speed of light be constant for all observers irrespective of their motion relative to the light's origin. In special relativity, the *principle of relativity* says that the laws of physics be the same for all objects moving in inertial reference frames (i.e., those of constant velocity). In general relativity, the *principle*

by means of special laws such as Newton's law of gravitation (or later Einstein's field equations)—knowledge that serves in the prediction of natural phenomena. The a priori principles of Newton's *absolute space* and *absolute time*, as Friedman stresses, changed in the transition from Newtonian to relativistic physics at the turn of the twentieth century, enabled by prior historical developments in mathematics and geometry and culminating in Einstein's theory of relativity. Following Hans Reichenbach (1920), who called them "axioms of coordination" (see Padovani 2011), Friedman treats such principles as *methodological* a priori propositions. Serving the same function as they did for Kant, the principles remain constitutive of experience but are not forever fixed, and so can be altered as science develops. Relativized to historical eras, then, a principle comes to reflect the systematic stage of scientific theorizing reached at a historical moment, rather than an inbuilt or otherwise pre-determined constraint on the cognitive apparatus of theory-users.

Once relativized, Friedman argues, a Neo-Kantian version of the a priori may be retained, while Kant's original one must be abandoned. This Neo-Kantian version is assigned with a significant task in the explanation of radical conceptual change in the transition from Newtonian to Einsteinian physics.[9] According to him, the explicit acknowledgement of philosophical meta-paradigms—including Kant's own, in which a priori forms of thinking have their modern historical origin—is necessary and, in combination with other elements of Friedman's account, also sufficient to successfully address the "Kuhnian challenge." The challenge, according to Friedman, is to make understandable how a rational form of mutual inter-paradigmatic communication between proponents of rivaling scientific paradigms, old and new, can be achieved when these paradigms are *incommensurable* in the sense of relying on non-intertranslatable conceptual frameworks, as explained above.

Friedman's response to Kuhn's challenge particularly arises from the need "to explain, prospectively, how the new framework becomes rational, a 'live' option" (Kindi 2011, 337) for those who adhere to the old paradigm. This notion of prospective rationality, holds Friedman, is distinct from the rationality afforded by a standard account of *inter-theory reduction* (see Batterman 2012). Here, the set of models, M, that are provided by a theory of the old paradigm, T, are (approximately) reduced to limiting cases of particular models from the set M* provided by a theory T* that belongs to the new paradigm. The models of T are thus

of equivalence expresses that "bodies affected only by gravitation follow geodesics in a variably curved space-time geometry" (Friedman 2002a, 187).

[9] "[W]hat I call the dynamics of reason is not intended to be a general theory of scientific change at all—rather, it is a particular historical narrative accompanied by a particular philosophical gloss. The point, on my view, is that the transition from Newton to Einstein is the most important challenge to the Enlightenment ideal of scientific rationality championed by Kant . . . , and I am attempting to respond to this challenge, accordingly, by reexamining this particular transition in considerably more historical detail." (Friedman 2010, 714)

(approximately) likened to those of T*.[10] For Friedman, however, a convergence of mathematical structures through inter-theory reduction can at most demonstrate that a paradigm shift is *retrospectively rational*. That is, the change from the old to the new theory is but unilaterally rational, namely from the point of view of the new paradigm. From the point of view of the old paradigm, however, an analogue to inter-theory reduction, one that would similarly render the transition to the new paradigm as a rational instance of conceptual development, is unavailable. Moreover, ontological or referential differences, as we have seen, provide Kuhn with the primary reason to attest a divergence of meaning between (the terms shared by) incommensurable conceptual frameworks.[11] And such referential meaning divergence remains unaddressed by limiting case reduction. Friedman essentially accepts referential divergence as capturing "a centrally important aspect of what he [Kuhn] has called the non-intertranslatability or 'incommensurability' of pre-revolutionary and post-revolutionary theories" (Friedman 2002a, 186, n.14). In particular, that such referential divergences arise in paradigm shifts would seem to make it more understandable that proponents of the old and the new paradigm face communicative difficulties—for they may find themselves "living in different worlds," worlds populated by different things—and so would render it more understandable that they allegedly have to *sway* one another into the adoption of one or the other paradigm.

14.5.2 The Role of Philosophical Meta-frameworks

Friedman's form of Neo-Kantianism takes issue with the holistic picture, standardly ascribed—as the Duhem-Quine thesis—to Pierre Duhem and Willard Van Orman Quine (see Brenner 1990), according to which our "web of belief" faces the tribunal of experience *in toto* (Duhem), so that each of its elements is as prone to modification as any other (Quine). Friedman instead proposes "an alternative picture of a thoroughly dynamical yet nonetheless differentiated system of knowledge that can be analyzed ... into three main components or levels" (Friedman 2002a, 189). The first level houses the component that faces experience directly, namely the

[10] In relativistic contexts, for instance, *momentum* is expressed as $p = m_0 v/\sqrt{(1-(v/c)^2)}$, where v is the velocity and c the speed of light. The special relativity form converges to the Newtonian $p = mv$ as v goes to zero. So the relativistic form reduces the Newtonian one, because $\sqrt{(1-(v/c)^2)}$ approaches 1 as $(v/c)^2$ approaches 0.

[11] Friedman places Kuhn within Cassirer's *genetic conception* that sees scientific knowledge to "progress from naively realistic 'substantialistic' conceptions, focusing on underlying substances, causes, and mechanisms subsisting behind the observable phenomena, to increasingly abstract purely 'functional' conceptions" (Friedman 2008, 244). But both Kuhn's notion of paradigm shift in *Structure* and his subsequent replacement for paradigm, the *structured lexicon*, also draw on diverging ontologies, or referent-shifts. Kuhn's position, Friedman argues, therefore also reflects the Meyersonian substantialistic view, although it is directly opposed to Cassirer's.

"empirical laws of nature, such as the Newtonian law of gravitation or Einstein's equations for the gravitational field" (ibid.). As we have seen, in our diction such laws or equations correspond to specific constraints on the distribution of points over a conceptual space.

The second level comprises the "constitutively a priori principles that define the fundamental spatio-temporal framework within which alone the rigorous formulation of first or base level principles is possible" (ibid.). From our perspective, second level principles correspond to (parts of) the conceptual space itself. For instance, principles expressing Newton's integral 3D space with its Euclidian metric, as well as Newton's separate 1D time or Einstein's integral 4D space-time with its non-Euclidian (Schwarzschild) metric explicitly count as a priori on the second level for Friedman. In contrast, Newton's separate 1D mass and the integral 3D force—being magnitudes that to successfully apply in predicting natural phenomena presupposes having assigned values on the space and time dimensions—may appear to count as being a priori at most implicitly. But Friedman goes on to claim that "[t]hese relativized a priori principles [at level two] constitute what Kuhn calls paradigms: relatively stable sets of rules of the game, as it were, that make possible the problem-solving activities of normal science [all of which involve the adoption, maintenance, or rejection of laws at level one]..." (ibid.), thus indicating that principles governing mass and force might count as a priori, too. After all, in order to predict a gravitational phenomenon such as a planetary orbit, for instance, one of the Newtonian "rules" of normal science prescribes that one enrich models that already feature values on the time and the space dimensions by stipulating suitable values for the dimensions of mass and force.[12] Friedman's second level would then comprise the entire conceptual framework or, in our diction, the conceptual space spanned by the dimensions with their metrics as these combine into domains.[13] This excludes from level two only the constraints on the distribution of points otherwise known as empirical laws, and located at his level one.

The second level provides Friedman with the necessary background so that the adoption, maintenance, and rejection of laws at level one—in the course of Kuhn's "normal science"—can be understood as an empirical matter, that is, as a matter

[12]In structuralist terms, the process is that of enriching a partial potential model of Newtonian mechanics to a full model of the theory. The structuralist approach is compared to the conceptual spaces approach in Gärdenfors and Zenker (2011) and Zenker and Gärdenfors (2014) where it is shown, among others, that the structuralist's three kinds of models—potential, partial potential, and full model—can be provided with spatial analogues. In each case, respectively, these models are understood as ever more restrictive constraints on the distribution of points over the space.

[13]Friedman perhaps comes closest to such distinctions when he reports the early Carnap as having offered "a generalization of Kant's conception of spatial intuition according to which only the infinitesimally Euclidean character of physical space is a priori determined by the form of our intuition ..., whereas the choice of specifically 'metrical form' (whether Euclidean or non-Euclidean) is 'optional [*wahlfrei*]'" (2002b, 24). Compare our second type of change (Sect. 14.3).

of methodologically-hardened[14] corroboration or falsification. On the other hand, the adoption, maintenance, or rejection of level two conceptual frameworks or paradigms—in the course of Kuhn's "revolutionary science"—cannot be handled as empirical processes in this sense. Friedman writes that,

> "no straightforward process of empirical testing [of the paradigm-constitutive a priori principles at level two], in periods of deep conceptual revolution, is then possible. Here our third level, that of philosophical meta-paradigms or meta-frameworks, plays and indispensable role, by serving as a source of guidance and orientation in motivating and sustaining the transition from one paradigm or conceptual framework to another. Such philosophical meta-frameworks contribute to the rationality of revolutionary change, more specifically, by providing a basis for mutual communication (and thus for communicative rationality in Habermas's sense) between otherwise incommensurable (and therefore non-intertranslatable) scientific paradigms." (Friedman 2002a, 189)

The role that Friedman envisions for philosophical meta-paradigms or meta-frameworks is that of "guiding the articulation of the new space of possibilities [delivered by the successor paradigm] and making the serious consideration of the new paradigm [by those committed to the old one] a rational and responsible option" (Friedman 2002a, 190). Like the second level, also the third functions in the spirit of a Kantian program, a program that seeks to explicate *conditions of possibility*. Among these meta-paradigms, for instance, Friedman counts the "new approach to the understanding of nature self-consciously crafted by Descartes and Galileo against the backdrop of medieval Scholasticism" (Friedman 2001, 23), and of course Einstein's "recognition of a new item, as it were, in the space of intellectual possibilities: namely the possibility of a relativized conception of time and simultaneity" (ibid.). A meta-paradigm shall deliver "new conceptions of what a coherent rational understanding of nature might amount to" (ibid.). It is, in brief, "a source of new ideas, alternative programs, and expanded possibilities that is not itself scientific in the same sense—that does not, as do the sciences themselves, operate within a generally agreed upon framework of taken for granted rules" (ibid.)

To summarize, the meta-paradigms at level three make possible the adoption, maintenance, and revision of a priori principles at level two; and the level two a priori principles make possible the adoption, maintenance, and revision of empirical laws at level one.

Friedman's position is thus that philosophical discourse is necessary for communicatively rational scientific progress in the face of radical conceptual change. The

[14]In criticism of Karl Popper's falsificationism, Imre Lakatos (1978) introduced the notion of a *methodologically hardened fact*. Unlike *naive falsificationists*, who treat empirical data ("facts") as indubitable—so that facts always "win" in case of a conflict between theory and observation—and also unlike *methodological falsificationists*, who treat as indubitable the observational theory that yields these facts, *sophisticated methodological falsificationists* "harden" their facts by presupposing a hierarchy of observational theories. So it is only the acceptance of an order of auxiliary theories that enables the theory to "lose" against experience (*falsification*) or to confirm it (*corroboration*). Conversely, when the predicting theory shall be maintained for some reason, then the facts may be "softened" by doubting the observational theory on which they rely (see Zenker 2009, ch. 4).

third level, in particular, is seen as *indispensable* in this process. So a consequence of Friedman's view seems to be this: when scientists each prefer (radically) different conceptual frameworks, but nevertheless manage to engage in rational communication about them, then they must engage, and must also to some extent be well-versed, in meta-philosophical discourse. We want to show that an alternative response to Kuhn's challenge is possible. One reason to look for an alternative response is that there are socio-empirical reasons to doubt Friedman's account: by and large, scientists are not well-versed in philosophical discourse, and it is not clear either that they need to be. If so, then the scientists' interaction in periods of radical theory change would for the most part fail to be communicatively rational in the sense that Friedman adopts from Habermas, or successful rational communication about diverging conceptual frameworks can instead be facilitated by something other than meta-philosophical discourse.

We suggest that the theory of conceptual spaces explains the communicative success, without committing to the third level of Friedman's three-tiered neo-Kantian view. Our aim in the next section, then, will be to argue that it is not necessary to deploy the entirety of philosophical discourse in order to explain that radical changes to a conceptual framework can be communicatively rational. If we are right, then the theory of conceptual spaces provides an alternative response to the Kuhnian challenge.[15]

[15] As this manuscript goes to print, we have become aware of a paper by Marta Sznajder (2014), in which she raises the objection that our invoking the theory of conceptual spaces in the present context does itself amount to an instance of meta-philosophical discourse. As she points out, moreover, Michael Friedman has provided historical evidence that particularly Isaac Newton and Albert Einstein in fact were aware of, and were influence by, the philosophical discourse at the time when they developed, and then proposed, what many of their contemporaries perceived as radically different new conceptual frameworks. If her objection holds, then our argument in the following section would presumably not suffice to establish that Friedman's third level is not necessary. Pending a more complete response, we here only remark that the objection does not so much undermine the claim we have raised, namely that conceptual spaces can serve in providing an account of how rational communication is nonetheless possible among scientists who each prefer a different paradigm, yet fail to command pronounced meta-philosophical abilities. Rather, this objection is directed at the related claim—which we do not wish to raise—that the new paradigm *is rational because* it has been established as a new option in ways that the application of conceptual spaces leave as communicatively rational interactions. After all, the term 'rational' has now taken on a stronger sense, for the latter claim relates to aspects that we view as going beyond the rationality of the communicative process through which the new paradigm is established as a hitherto unavailable option, as laid out above. This other claim, thus, pertains to the new paradigm's 'ultimate rationality', that is, whether the new paradigm is *all-things-considered* better than the old paradigm. To adequately address this question presumably requires things other than conceptual spaces alone, and would seem to become clearer only as the new paradigm "proves its mettle" by applying it for the kinds of predictive and technological purposes that empirical theories tend to serve.

14.6 Communication About Theoretical Frameworks as Conceptual Spaces

It is crucial for our argument to recognize that, like Kuhn's (1970, 1987, 2000), also Friedman's account remains grounded in a tradition that understands conceptual frameworks as *linguistic entities*, while ours models these entities as (abstract) spaces. As we have seen, one may understand genuinely empirical laws as symbolic expressions of constraints on the distribution of points over the space spanned by its dimensions. This was perhaps clearest for cases such as Newton's law of gravitation that predicts these points to lie on the hyper-surface described by $F = GMm/r^2$ (Gärdenfors and Zenker 2013). Recall further that the law restricts these points to a subspace of that described by Newton's second law, $F = ma$, which is normally treated as an axiom of the theory. In brief, on our account a theory consists of a conceptual space (the framework) together with constraints on empirically possible points in the space (that can be expressed as laws).

This spatial view extends also to propositions such as Einstein's light principle that, for Friedman, has a priori status (see DiSalle 2002, 196f.). To illustrate, the principle says that the speed of light, c, is constant in all inertial reference frames, where c can be dimensionally analyzed as $[LT^{-1}]$, i.e., a length [L] over a time interval [T]. Interpreted in terms of conceptual spaces, the light principle restricts c to a single value, given by the quotient of the dimensions [L] and [T], of 299,792,458 m/s. In relativistic treatments of 4D space-time, moreover, 3D space and 1D time are integral dimensions and so form a domain. This contrasts with the Newtonian theory where 3D space and 1D time constitute separate domains, gravity is a force acting-at-a-distance, and the speed of light an *additive* magnitude (rather than bound to c). And it similarly contrasts with a Cartesian view where, being unrestricted, c marks infinity—giving rise to an interpretation of light signals as propagating instantaneously, and so coordinating distant events to local ones.

When a conceptual framework is analyzed as a spatial structure, the empirical laws at Friedman's first level are demoted in status because, as we have seen, they are merely the linguistic means to express constraints on a conceptual space. And the same holds for the linguistic expressions known as a priori principles or axioms at the second level. After all, the information that linguistic approaches take to be expressed by such principles may partly be read off directly from the structure of the space, and so is generated by the combination of dimensions and their metric.[16] And it can partly be read off by interpreting such principles to constrain the space, so that predictions warranted by a theory's empirical content fall within the space's hyper-planes. Consequently, a new conceptual framework F*,

[16]To give an elementary example, the fact that relations on dimension such as "longer than" and "warmer than" are *transitive* follows immediately from their one-dimensional structure, being isomorphic with the real line. Thus the transitivity need not be formulated as an axiom, as is standard in most linguistic or logical accounts of theories, but is inherent in the dimensional structure of the underlying conceptual space.

with its embedded theory T*, that *arise together* as a new paradigm in the course of Kuhn's revolutionary science, can be described by applying the five change operations (Sect. 14.3) to the predecessor framework F—both F and F* having been reconstituted as conceptual spaces. And the set of models comprising the empirical content of the new theory T*, that is formulated against the background of F*, may similarly be understood as a distinct set of constraints.

The linguistic approach that is used by both Kuhn and Friedman identifies such changes in terms of a replacement of sentences, rather than describing the changes to a priori principles, axioms, and empirical laws of T into those of T* that take place in the transition from the conceptual framework F to the successor F* as being *determined* by a structural modification of F. Furthermore, there is no evidence in Friedman's writings of a principled distinction between a conceptual framework, on the one hand, and an empirical theory that is expressed relative to that framework, on the other. (In our model, theories are defined by *constraints* on the conceptual space of a theory framework.) For instance, Friedman writes that

> "[t]he key move in general relativity ... is to *replace* the law of inertia—which, from the space-time perspective inaugurated by special relativity, depicts the trajectories of force-free bodies as geodesics in a flat space-time geometry—with the principle of equivalence, according to which bodies affected only by gravitation follow geodesics in a variably curved space-time geometry." (Friedman 2002a, 187, *italics added*)

We take such passages as evidence that both the empirical theory and the conceptual framework that it presupposes are primarily understood as linguistic entities, so that an analysis of their dynamics will need to rely on change operations like those described in the AGM tradition (Alchourrón et al. 1985; Gärdenfors 1988).[17] Here, for instance, the replacement or revision of proposition p by q is modeled as the retraction of p followed by the addition of q; the various propositions of a theory, moreover, are ordered according to levels of entrenchment (Zenker 2009).

We next turn to the consequences of adopting the spatial perspective on theories and their frameworks for scientific communication. As we have seen, a model of T is a constrained hyperspace of F which itself is a spatial object constituted by dimensions, their metrics, and the way these dimensions form domains. Such spatial objects, we submit, form the basis for communication between scientists during periods of what Friedman calls "deep conceptual change." These spatial

[17] We take Friedman's employment of the terms "succession," "transformations" and "extension" to reflect the same understanding:

> "I have argued, on the one hand, that the transition from Newton to Einstein centrally involves a succession of relativized constitutively a priori principles ..., and the existence of such diverse constitutively a priori principles, on my view, captures the essence of Kuhnian incommensurability. But I have also suggested, on the other hand, that the detailed historical route from earlier to later constitutive principles exhibits the latter as natural transformations of the former—arising as a sequence of 'minimal extensions' of our Kantian-Newtonian starting point in a succession of new mathematical, empirical, and philosophical situations." (Friedman 2010, 713)

objects allow for identifying some or another set of dimensions and constraints as being methodologically a priori—making it expectable that scientists display diverging preferences as to which dimensions in fact are fundamental in this sense—because none of these dimensions count as inherently privileged. For instance, in electrodynamics, particle theorists tend to view *electrical current* as a fundamental dimension, while field theorists take *electrical charge* as fundamental (Gärdenfors and Zenker 2013). And just like Joseph Sneed's (1971) distinction between T-theoretical and T-non theoretical terms, on which structuralists rely (Gärdenfors and Zenker 2011), also the Neo-Kantian distinction between a priori principles and empirical laws remains perfectly possible in the context of conceptual spaces.

Our critical objection, however, is that neither such distinctions nor intimate knowledge of the philosophical traditions from which they arise appear to be *indispensable* for communication to remain rational in the relevant sense. On a cognitive account of meaning and a representation of theories with the aid of conceptual spaces, rather, communicative rationality appears to be available prior to Friedman's third level. So, one *may* assign a priori status to some but not other dimensions of an empirical theory's conceptual framework. But the same appears not to be necessary in order to explain that rational communication between proponents of different paradigms, about these differences, is possible during periods of deep conceptual change.

Let us further spell out our position. We do not claim that, when successfully communicating, scientists consciously entertain the theory of conceptual spaces as a shared background tool. Nor do we claim that, if scientists were to entertain—consciously or not—the theory of conceptual spaces, they would never disagree about issues of theory choice for, as it were, the need to persuade, and be persuaded, never arises. Rather, our claim is that the theory of conceptual spaces can explain scientists' communicative successes—particularly those related to issues arising inter-paradigmatically—without postulating meta-paradigmatic philosophical abilities on their behalf. Our view moreover entails that persuasive attempts at bringing an opponent to adopt a scientific framework she does not prefer need not, *ab initio*, be outside the realm of rationally reconstructable discourse. This is in contrast to the Kuhnian view we criticize.

At the same time, our account offers little in the way of a positive characterization of, or general guidance on, how scientists *should* rationally persuade one another. Analysts studying actual scientific discourse may nevertheless find it worth considering that inter- and intra-paradigmatic communication failure and success alike are not readily explainable by citing the discourse participants' differentially pronounced meta-philosophical abilities. Applying the theory of conceptual spaces also for the purpose of describing scientific discourse will make it easier to provide good accounts of the communicative phenomena.

14.7 Conclusion

Philosophy of science has put itself in a deadlock by representing theories as linguistic entities (symbolically expressed systems of laws) and by assuming semantic realism. This leads to problems in understanding how scientists can communicate rationally in times of radical theory change. The deadlock witnessed, among others, Kuhn's claim that paradigms are incommensurable. This claim ultimately accepts, as an *explanandum*, what in fact is a reconstructively incurred artifact owed to Kuhn's own theoretical perspective, from which theories are seen as linguistic or symbolic structures. As a remedy, Friedman introduces meta-philosophical discussions as a way of accounting for how scientists can communicate in a rational way. However, scientists would generally not see the need for such discussions to resolve the conflict between two competing theories, even if the frameworks of the theories are different.

In contrast, if theory frameworks are modeled as conceptual spaces and theories as constraints on such spaces, the communicative challenge will become smaller. Frameworks have to be compared in scientific discussions, but this can be handled by comparing their geometric or topological properties, instead of relying on meta-philosophical considerations. And once theory frameworks are thus comparable, the incommensurability aspect of conceptual change collapses into referential divergence, or ontological change. In brief, a discussion conducted in terms of meta-philosophical principles won't add much towards guiding "the articulation of the new space of possibilities" and making "the serious consideration of the new paradigm a rational and responsible option" (Friedman 2002a, 190).

We have provided reasons to doubt that addressing radical meaning changes in adequate ways requires, or makes indispensable, the meta-philosophical principles at the third level of Friedman's account. Instead, we have argued, if mutual inter-paradigmatic communication between scientists shall be saved—so that, also from the point of view of the old paradigm, the new conceptual framework can be seen as rationally superior to the former (Kindi 2011, 337)—then a reconstruction of the successive conceptual frameworks as two conceptual spaces already allows for such mutual inter-paradigmatic communication. So Friedman's three-tiered method of handling Kuhn's challenge can be replaced by a cognitively and communicatively more economical account.

In conclusion, we find that Friedman's neo-Kantian perspective has some advantages over a sentence-based account (such as logical positivism) or Kuhn's position. Nevertheless, his third level has little role to play in actual scientific discourse; for most purposes that are of concern to scientists, it is sufficient to rely on a comparison between the spatial frameworks of the theories under scrutiny.

Acknowledgements Previous versions of this paper were presented at the 2014 Meeting of the Nordic Network for Philosophy of Science, 27–28 March 2014, and the "Conceptual Spaces at Work" conference, 24–26 May 2012, both held at Lund University, Sweden. For discussion and useful comments, we thank audience members and two anonymous reviewers. Both authors acknowledge funding from the Swedish Research Council.

References

Alchourrón, C., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic, 50*, 510–530.

Batterman, R. (2012). Intertheory relations in physics. In E. N. Zalta, (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2012 Edition). http://plato.stanford.edu/archives/fall2012/entries/physics-interrelate. Accessed 23 Jan 2014

Brenner, A. A. (1990). Holism a century ago. The elaboration of Duhem's thesis. *Synthese, 83*, 325–335.

D'Agostino, F. (2013). Verballed? Incommensurability 50 years on. *Synthese, 191*(3), 517–538.

DiSalle, R. (2002). Reconsidering Kant, Friedman, logical positivism, and the exact sciences. *Philosophy of Science, 69*, 192–211.

Einstein, A. (1924). Review of Elsbach 1924. *Deutsche Literaturzeitung, 45*, 1685–1692.

Friedman, M. (1998). On the sociology of scientific knowledge and its philosophical agenda. *Studies in History and Philosophy of Science, 29*(2), 239–271.

Friedman, M. (2001). *Dynamics of reason*. Stanford: CSLI Publications.

Friedman, M. (2002a). Kant, Kuhn, and the rationality of science. *Philosophy of Science, 69*, 171–190.

Friedman, M. (2002b). Kuhn and logical empiricism. In T. Nickles (Ed.), *Thomas Kuhn* (pp. 19–44). Cambridge, UK: Cambridge University Press.

Friedman, M. (2008). Ernst Cassirer and Thomas Kuhn: The neo-Kantian tradition in history and philosophy of science. *Philosophical Forum, 39*, 239–252.

Friedman, M. (2010). Synthetic history reconsidered. In M. Domski & M. Dickson (Eds.), *Discourse on a new method. Reinvigorating the marriage of history and philosophy of science* (pp. 571–813). Chicago/La Salle: Open Court.

Gärdenfors, P. (1988). *Knowledge in flux*. Boston: MIT Press.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press.

Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. Cambridge, MA: MIT Press.

Gärdenfors, P., & Zenker, F. (2011). Using conceptual spaces to model the dynamics of empirical theories. In E. J. Olsson & S. Enqvist (Eds.), *Philosophy of science meets belief revision theory* (pp. 137–153). Berlin: Springer.

Gärdenfors, P., & Zenker, F. (2013). Theory change as dimensional change: Conceptual spaces applied to the dynamics of empirical theories. *Synthese, 190*(6), 1039–1058.

Habermas, J. (1984) [1981]. *Theory of communicative action volume one: Reason and the rationalization of society* (T. A. McCarthy, Trans.). Boston: Beacon Press.

Habermas, J. (1987) [1981]. *Theory of communicative action volume two: Liveworld and system: A critique of functionalist reason* (T. A. McCarthy, Trans.). Boston: Beacon Press.

Howard, D. (2010). 'Let me briefly indicate why I do not find this standpoint natural.' Einstein, general relativity, and the contingent a priori. In M. Domski & M. Dickson (Eds.), *Discourse on a new method. Reinvigorating the marriage of history and philosophy of science* (pp. 333–355). Chicago/La Salle: Open Court.

Kindi, V. (2011). The challenge of scientific revolutions: Van Fraassen's and Friedman's responses. *International Studies in the Philosophy of Science, 25*(4), 327–349.

Kuhn, T. (1970 [1962]). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

Kuhn, T. (1987). What are scientific revolutions? In L. Krüger, L. Daston, & M. Heidelberger (Eds.), *The probabilistic revolution volume one* (pp. 7–22). Cambridge, MA: MIT Press.

Kuhn, T. (2000). *The road since structure. Philosophical essays, 1970–1993* (with an autobiographical interview, J. Conant & J. Haughland, Ed.), Chicago: University of Chicago Press.

Lakatos, I. (1978). *The methodology of scientific research programs*. Cambridge, UK: Cambridge University Press.

Ladyman, J. (2014). Structural realism. In E.N. Zalta, (Ed.), *The Stanford encyclopedia of philosophy*. http://plato.stanford.edu/archives/spr2014/entries/structural-realism/. Accessed 22 Nov 2014

Padovani, F. (2011). Relativizing the relativized a priori: Reichenbach's axioms of coordination divided. *Synthese, 181*(1), 41–62.

Petersen, G., & Zenker, F. (2014). From Euler to Navier-Stokes: A spatial analysis of conceptual changes in 19th-century fluid dynamics. *International Studies in the Philosophy of Science, 28*(3), 1–19.

Rehg, W. (2009). *Cogent science in context. The science wars, argumentation theory, and Habermas*. Cambridge, MA: MIT Press.

Reichenbach, H. (1920). *Relativitätstheorie und Erkenntnis apriori*. Berlin: Springer. M. Reichenbach, Trans. (1965). *The theory of relativity and a priori knowledge*. Berkeley/Los Angeles: University of California Press.

Reisch, G. A. (1991). Did Kuhn kill logical empiricism? *Philosophy of Science, 58*(2), 264–277.

Schilpp, P. A. (Ed.). (1963). *The philosophy of Rudolf Carnap*. La Salle: Open Court.

Sneed, J. D. (1971). *The logical structure of mathematical physics*. Dordrecht: Reidel.

Sznajder, M. (2014). Inductive logic and conceptual spaces: Carnap's basic system and beyond. Draft available from the author.

Zenker, F. (2009). *Ceteris Paribus in conservative belief revision*. Berlin: Peter Lang.

Zenker, F. (2014). From features via frames to spaces. Modeling scientific conceptual change without incommensurability or aprioricity. In T. Gamerschlag, D. Gerland, R. Osswald, & W. Petersen (Eds.), *Frames and concept types: Applications in language and philosophy* (pp. 69–89). Dordrecht: Kluwer/Reidel.

Zenker, F., & Gärdenfors, P. (2014). Modeling diachronic changes in structuralism and in conceptual spaces. *Erkenntnis, 79*(8), 1547–1561.