Lucas Bechberger
Kai-Uwe Kühnberger
Mingya Liu *Editors*

Concepts in Action

Representation, Learning, and Application

Springer

Language, Cognition, and Mind

Volume 9

This series takes the current thinking on topics in linguistics from the theoretical level to validation through empirical and experimental research. The volumes published offer insights on research that combines linguistic perspectives from recently emerging experimental semantics and pragmatics as well as experimental syntax, phonology, and cross-linguistic psycholinguistics with cognitive science perspectives on linguistics, psychology, philosophy, artificial intelligence and neuroscience, and research into the mind, using all the various technical and critical methods available. The series also publishes cross-linguistic, cross-cultural studies that focus on finding variations and universals with cognitive validity. The peer reviewed edited volumes and monographs in this series inform the reader of the advances made through empirical and experimental research in the language-related cognitive science disciplines.

For inquiries and submission of proposals authors can contact the Series Editor, Chungmin Lee at chungminlee55@gmail.com.

More information about this series at http://www.springer.com/series/13376

Lucas Bechberger · Kai-Uwe Kühnberger ·
Mingya Liu
Editors

Concepts in Action

Representation, Learning, and Application

Editors
Lucas Bechberger
Institute of Cognitive Science
Osnabrück University
Osnabrück, Germany

Kai-Uwe Kühnberger
Institute of Cognitive Science
Osnabrück University
Osnabrück, Germany

Mingya Liu
Department of English and American
Studies
Humboldt University of Berlin
Berlin, Germany

Acknowledgements

Contents

Contributors

Lucas Bechberger Institute of Cognitive Science, Osnabrück University, Osnabrück, Germany

Michael Färber Karlsruhe Institute of Technology (KIT), Institute AIFB, Karlsruhe, Germany

Paola Gega Institute of Philosophy, University of Bochum, Bochum, Germany

Helmar Gust Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

Andreas Harth Friedrich-Alexander-University Erlangen-Nuremberg, Nuremberg, Germany;
Fraunhofer IIS-SCS, Nuremberg, Germany

Cristina Iani Department of Surgery, Medicine, Dentistry and Morphological Sciences with Interest in Transplant, Oncology and Regenerative Medicine, University of Modena and Reggio Emilia, Reggio Emilia, Italy;
Center for Neuroscience and Neurotechnology, University of Modena and Reggio Emilia, Modena, Italy

Kai-Uwe Kühnberger Institute of Cognitive Science, Osnabrück University, Osnabrück, Germany

Mingya Liu Department of English and American Studies, Humboldt University of Berlin, Berlin, Germany

Andreas Nürnberger Data and Knowledge Engineering Group, Faculty of Computer Science, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany

Sandro Rubichi Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Reggio Emilia, Italy;
Center for Neuroscience and Neurotechnology, University of Modena and Reggio Emilia, Modena, Italy

Elisa Scerrati   Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Reggio Emilia, Italy

Stefan Schneider   Data and Knowledge Engineering Group, Faculty of Computer Science, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany

Yulia Svetashova   Robert Bosch GmbH, Corporate Research and Advance Engineering, Renningen, Germany

Carla Umbach   Department of German Language & Literature I, University of Cologne, Cologne, Germany

Paola Vernillo   Università degli Studi di Firenze (UNIFI), Firenze, Italy

Concepts in Action: Introduction

Lucas Bechberger and **Mingya Liu**

It is impossible to talk about human cognition without talking about concepts—there simply *is* no human cognition without concepts. Concepts form an abstraction of reality that is central to the functioning of the human mind. Conceptual knowledge (of e.g., APPLE, LOVE and BEFORE) is crucial for us to categorize, understand, and reason about the world. Only equipped with concepts and words for them can we successfully communicate and carry out actions. *But what exactly are concepts? How are concepts acquired? How does the human mind use concepts?* Such questions have been a subject of discussion since antiquity and remain highly relevant in multiple fields (e.g., Murphy 2002; Margolis and Laurence 2015).

Recent decades have seen fruitful results and methodological advances on concept research in disciplines such as linguistics, philosophy, psychology, artificial intelligence, and computer science. For instance, cognitive psychologists use empirical experiments to validate formal models of concept representation and learning such as the prototype theory (Rosch et al. 1976), the exemplar theory (Murphy 2016) or other alternative theories (Rogers and McClelland 2004; Blouw et al. 2016). Linguists pursue the goal of assigning more precise meaning to natural language expressions by mainly applying logic-based formalisms (Asher 2011). In machine learning, decision boundaries in high-dimensional feature spaces are used to define membership to a concept (Mitchell 1997). Moreover, researchers in the semantic web area have

L. Bechberger (✉)
Institute of Cognitive Science, Osnabrück University, Osnabrück, Germany
e-mail: lucas.bechberger@uni-osnabrueck.de

M. Liu (✉)
Department of English and American Studies, Humboldt University of Berlin, Berlin, Germany
e-mail: mingya.liu@hu-berlin.de

created large ontologies (Gómez-Pérez et al. 2004) containing hierarchies of concepts formulated in description logics. Google's "Knowledge Graph" illustrates how such ontologies can be used in industrial applications.

Despite of this plethora of research, there remain many open questions, unsolved debates and methodological challenges. For instance, the ontologies of the semantic web have been challenged as being unable to represent information about conceptual similarity and thus as being ill-suited for representing conceptual knowledge (Gärdenfors 2004). And although deep learning models are often said to acquire concepts when learning to classify pictures of dogs, umbrellas, and other objects, they can be easily fooled by slightly manipulated input images (Szegedy et al. 2013)—which highlights that they only learn patterns, but no conceptual knowledge.

One major obstacle for a better and more holistic understanding of concepts is that research on concepts has usually been carried out in different disciplines individually—with different approaches, different goals, and different results. The multi-disciplinary research efforts usually run in parallel without enough interaction; existing interdisciplinary research projects usually do not involve more than two disciplines, for example linguistics and computer science in the WordNet project (Fellbaum and Vossen 2016) or psychology and artificial intelligence in cognitive architectures like ACT-R (Anderson 2009) or SOAR (Laird 2012). In order to move the scientific understanding of concepts forward, we need a truly interdisciplinary perspective on concepts, involving a mutual understanding of the different approaches from different disciplines, a lively exchange of ideas, and synergies arising from the combination of different research perspectives and methods. Thus, our volume will focus on selected recent issues, approaches and results that are not only central to the highly interdisciplinary field of concept research, but that are also particularly important to newly emergent paradigms and challenges.

This volume focuses on three topics (i.e., three distinct points of view) that lie at the core of concept research: *representation*, *learning*, and *application*. In the following, we will first present research questions related to the three topics (Sect. 1), and then, we will provide an overview of the contributions (Sect. 2).

1 Research Questions

In order to structure an interdisciplinary discussion and exchange about concept research, we found it useful to put a focus on three essential questions that need to be answered: How can conceptual knowledge be represented (Sect. 1.1)? How are concepts acquired (Sect. 1.2)? How is conceptual knowledge applied in cognitive tasks (Sect. 1.3)?

1.1 Representation: How Can We Formally Describe and Model Concepts?

One of the major challenges in concept research is to find a formal representation of concepts that is on the one hand able to explain a wide range of empirical observations and experimental results and that can on the other hand be easily applied in practice. Exemplar and prototype theories from psychology focus on the crucial role of representative instances, whereas knowledge-based theories (Murphy and Medin 1985) emphasize that concepts do not occur in isolation, but always stand in relations to other concepts. Ontologies (Gómez-Pérez et al. 2004) from the semantic web area provide a formal way of describing such networks of concepts. The logical-formal approaches from linguistics aim at accounts of the context-independent and context-dependent aspects of meaning and can be related to logic-based representations in artificial intelligence (Russell and Norvig 2002). Finally, the feature spaces commonly used in the field of machine learning (Mitchell 1997) (for example in nearest-neighbor classifiers) can be linked to prototype and exemplar approaches from psychology. When analyzing formal representations of concepts, the following questions should be considered:

- What are the underlying assumptions of different representation approaches? How are they motivated and to what extent are they compatible with one another?
- How can different representation formalisms be compared with one another? What are useful and meaningful criteria for making such a comparison?
- How transferable are the different approaches to other domains? For instance, are there any benefits in using a prototype approach in linguistic analyses of natural language semantics?
- How can different representational approaches be augmented or combined with one another in order to arrive at a more holistic model?

1.2 Learning: Where Do Concepts Come from and How Are They Acquired?

Another major issue in concept research is concerned with concept acquisition, which is not only important per se but also essential for evaluating whether a specific theory of human concepts is psychologically plausible (Carey 2015). While there are well-established assumptions about children's acquisition of core concepts such as the basic-level bias and the taxonomic assumption, the exact nature of the underlying processes remains controversial. On a larger time scale, the evolution of concepts in human societies (Hull 1920) and similar processes in groups of robots (Spranger 2012) can give insights into learning processes. Moreover, studying concept learning across languages and cultures enables a better understanding of universality and diversity in concepts (Imai et al. 2010). Furthermore, to successfully coin and transfer new concepts, it is crucial to understand differences between everyday concepts and

expert concepts, e.g., in mathematics (Rips et al. 2008). These and other related issues (e.g., innateness, groundedness and embodiment) require researchers to not only strive for advances in their own field (such as in terms of improved machine learning algorithms in artificial intelligence), but also to start in-depth exchanges with neighboring disciplines. The following questions can provide useful guidelines when approaching concept learning:

- Which kinds of learning mechanisms (e.g., supervised vs. unsupervised, multi-modal vs. unimodal) are in the focus of the different research disciplines? Can multiple learning mechanisms be combined with one another?
- Which representational assumptions are made by the different learning mechanisms?
- How does concept learning interact with the development of low-level and other high-level cognitive (e.g., motorsensory, reasoning) abilities? What are their underlying mechanisms?
- What are the differences between learning concrete (e.g., APPLE) and abstract concepts (e.g., LOVE) and between learning expert and everyday concepts?

1.3 Application: How Are Concepts Used in Cognitive Tasks?

The last decade has witnessed an exploding utilization of conceptual knowledge bases, unprecedented both in scale and range of applications. The conceptual core of the semantic web (Berners-Lee et al. 2001) and artificial agents like IBM's Watson (Ferrucci et al. 2010) is largely based on AI technologies dating back to the last millennium (e.g., description logics). The new development has clearly shown the potential but also the limits of such approaches. The questions that arise here obviously link to other fields: The combination of a multitude of potential resources asks for modern AI methods to reason over heterogeneous and inconsistent data (Potyka and Thimm 2017). The application of conceptual knowledge in communication, including conceptual combination and application of conceptual knowledge in context, are classical problems in linguistics. And the problem of generating new concepts may find answers in recent psychological theories on creativity (Schorlemmer et al. 2014). The following questions are important with regard to the application of concepts:

- Which aspects of applying concepts are analyzed in the different disciplines? Is there any considerable overlap?
- How do the different representational formalisms constrain specific application scenarios and vice versa?
- Which contextual effects occur in the application of concepts? How are these effects handled in different frameworks?
- Which mechanisms exist for performing conceptual combination? Which constraints apply? How are conflicts that emerge from conflicting conceptualizations detected and resolved?

2 Summaries of the Contributed Chapters

This volume consists of seven individual chapters from different scientific disciplines, each of which relates to at least one of the other topics presented in Sect. 1. Figure 1 illustrates how the individual contributions relate to each other, based on their underlying disciplines, common themes, and the three focus topics from Sect. 1. Figure 1 illustrates both the strong relations between the individual contributions and the broad spectrum of this edited volume. We will now introduce the individual contributions in more detail.

Bechberger and Kühnberger's contribution "Generalizing Psychological Similarity Spaces to Unseen Stimuli – Combining Multidimensional Scaling with Artificial Neural Networks" (Chap. 2) addresses the focus topic of *learning*. It uses a *spatial model* of concepts as regions in psychological similarity spaces based on Gärdenfors' cognitive framework of conceptual spaces. These similarity spaces are typically obtained based on dissimilarity ratings from psychological studies and the technique of "multidimensional scaling" (MDS). This approach is however unable to generalize to unseen inputs. The authors propose to use MDS on human similarity ratings for initializing the similarity space and ANNs (artificial neural networks) to learn a mapping from raw stimuli into this similarity space. This proposal is a valuable contribution for integrating *psychology* and *artificial intelligence*. In order to
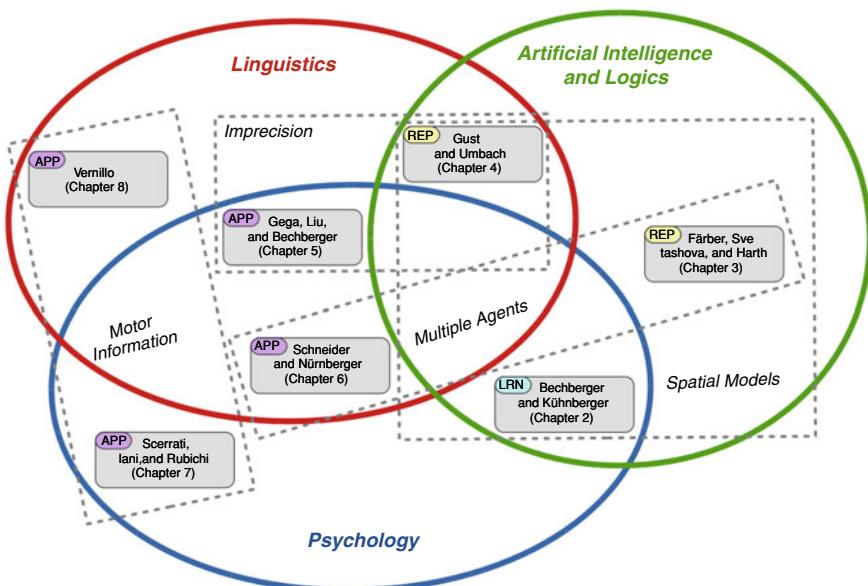
Fig. 1 Visualization of the contributed chapters based on scientific disciplines (solid ellipses), common research themes (dashed rectangles), and classification based on the three focus topics representation (REP), learning (LRN), and application (APP)

validate their hybrid approach, the authors conducted a feasibility study. Their results show that while their proposal works in principle, the generalization capabilities of the ANNs are still limited and need to be improved further.

Färber, Svetashova, and Harth's contribution "Theories of Meaning for the Internet of Things" (Chap. 3) is concerned with the *representation* of concepts in the context of the Internet of Things (IoT) from the perspective of *artificial intelligence*. They compare different representational frameworks from philosophy and computer science, taking a simple smart home setting as an application example. Overall, they consider four different approaches, namely model-theoretic semantics (which are based on first-order logic), possible world semantics (using modal logic), situation semantics, and cognitive and distributional semantics (i.e., *spatial models* of meaning). With the IoT application in mind, the authors assess whether these representational frameworks are able to represent intersubjectivity (i.e., *multiple agents*) and dynamics (i.e., changes in the state of the world) and to what extent they can be connected to perception. The authors conclude that none of the existing approaches is able to completely satisfy all three requirements. They propose to further investigate a combination between situational and distributional semantics as a promising avenue for future research.

Also **Gust and Umbach**'s contribution "A Qualitative Similarity Framework for the Interpretation of Natural Language Similarity Expressions" (Chap. 4) explores the *representation* of concepts in the context of natural language semantics. It aims at the interpretation of expressions of similarity and sameness, such as *so/similar/same* in English or their counterparts in German. The authors argue that treating similarity as a primitive predicate is unsatifactory because semantic differences between individual similarity expressions could not be accounted for and the role of similarity expressions in creating ad-hoc kinds, for example, by similarity demonstratives and scalar and non-scalar equatives would be obscured. The framework proposed in the paper introduces a non-metric qualitative concept of similarity which makes use of a *spatial model* called attribute spaces equipped with systems of predicates corresponding to predicates on the domain. Individuals are mapped to points in attribute spaces by generalized measure functions. Two individuals count as similar if their images in a particular attribute space given a particular predicate system cannot be distinguished. This allows representations of varying granularity and hence of different degrees of *imprecision*. The authors argue that the framework is suited for modeling the meaning of natural language similarity expressions and, moreover, account for their role in ad-hoc kind formation constructions. It thus provides a *logic*-based formalism which is able to explain *linguistic* phenomena.

Gega, Liu and Bechberger's contribution "Numerical Concepts in Context" (Chap. 5) deals with the semantics and pragmatics of numerical expressions, with a focus on their precise or *imprecise* interpretations. While the precise interpretation most prominently appears in mathematical contexts, the imprecise interpretation seems to arise when numbers (as quantities) are applied to real world contexts (e.g., *the rope is 50 m long*). Earlier literature shows that the (im)precise interpretation can depend on different factors, e.g., the kind of approximators a numeral appears with (precise vs. imprecise, e.g., *exactly* vs. *roughly*) or the kind of the numeral

itself (round vs. non-round, e.g., 50 vs. 47). The authors report on a *corpus-linguistic* study and a *psycholinguistic* rating experiment of English numerical expressions. The results confirm the effects of both factors, and additionally also an effect of the kind of unit, namely, whether it refers to discrete versus continuous concepts (e.g., PEOPLE vs. METER).

Schneider and Nürnberger's contribution "Evaluating Semantic CoCreation by using a Marker as Linguistic Constraint in Cognitive Representation Models" (Chap. 6) explores the *application* of conceptual knowledge in communication between *multiple agents*. More specifically, they address semantic co-creation, i.e., the convergence of the cognitive models of the interlocutors within a conversation. The authors hypothesize that a shared marker can facilitate this coordination of representations. In order to validate this hypothesis, they conducted an experiment where groups of three participants needed to identify a target location on a given map. One participant (the describer) was given the target and had to describe it to the two other participants who needed to correctly identify this target location. One of them (the committer) was able to give feedback to the describer while the other one (the observer) had to remain passive. The authors considered four experimental conditions which differed in the availability of a shared marker (i.e., a movable point on the map) and in the complexity of the task (measured by the number of cities displayed on the map). Their results show that when task complexity was low, no real interaction between the participants was necessary to successfully solve the task. Contrary to their expectations, the shared marker was not able to improve performance in the high-complexity scenario. While their results highlight that a certain level of complexity is necessary to elicit interactions, it also casts doubt on the assumption that additional means of communication (such as a shared marker) necessarily improve the outcome of the interaction. Their work thus urges for further research both in *psychology* and *linguistics* to gain a deeper understanding of the observed effects.

Scerrati, Iani and Rubichi's contribution "Does the Activation of Motor Information Affect Semantic Processing?" (Chap. 7) considers the *application* of concepts in lexical decision tasks, focusing on the influence of pre-activated *motor information*. The authors report on a *psychological* priming experiment in which the subjects were instructed to make keypress responses depending on two factors: One factor is word type with target words being relevant/irrelevant/unrelated to action (e.g., *handle/ceramic/eyelash*) with respect to a prime object (e.g., image of a frying pan). The other factor is spatial compatibility with the related part of the prime object (e.g., *handle* for a frying pan) either on the same side or on the opposite side of the key to be pressed. The dependent measures were reading time (RT) latencies and error rates for the question whether the target word was an Italian word. The results of the RT latencies did not show any significant effects or an interaction. The results of the error rates however showed a significant main effect of word type with the lexical decision responses being more accurate with action-relevant target words than with action-irrelevant words or unrelated words. This indicates that motor activation may indeed influence semantic processing, thus complementing and enriching the literature that focuses on the reverse effect of semantic content on motor activation.

Also **Vernillo**'s contribution "Grounding Abstract Concepts in Action: The Semantic Analysis of Four Italian Action Verbs Encoding Force Events" (Chap. 8) focuses on the *application* of conceptual knowledge, comparing the concrete and metaphorical uses of the four Italian action verbs *premere, spingere, tirare,* and *trascinare* (i.e., 'press', 'push', 'pull', and 'drag'). The underlying hypothesis is that the image schema of their literal meaning also constrains their usage in the metaphorical meaning. The *linguistic* study uses the representation of verb meanings through 3D scenes from the IMAGACT database. Based on the extracted data, the author provides a description of the semantic resemblances and differences in terms of salient image-schematic structures. The results show that the four verbs under consideration belong to the same semantic class of force (involving *motor information* and movement), and that they share commonalities in their literal and metaphorical use. At the same time, one can also observe systematic differences: For instance, while the literal meaning of *premere* focuses on the force exerted on the object, *spingere* emphasizes the resulting movement. These different connotations are also transferred to the metaphorical usage where *spingere* entails a change of state while *premere* does not. The results of this analysis support the view that metaphors are not just a linguistic phenomenon, but are grounded in embodied conceptual knowledge.

References

Anderson, J. R. (2009). *How can the human mind occur in the physical universe?* Oxford University Press.

Asher, N. (2011). *Lexical meaning in context: A web of words*. Cambridge University Press.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, *284*(5), 34–43.

Blouw, P., Solodkin, E., Thagard, P., & Eliasmith, C. (2016). Concepts as semantic pointers: A framework and computational model. *Cognitive Science*, *40*(5), 1128–1162.

Carey, S. (2015). Why theories of concepts should not ignore the problem of acquisition. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 415–454). MIT Press.

Fellbaum, C., & Vossen, P. (2016). *Proceedings of the 8th Global Wordnet Conference*.

Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A. A., et al. (2010). Building Watson: An overview of the DeepQA project. *AI Magazine*, *31*(3), 59–79.

Gärdenfors, P. (2004). How to make the semantic web more semantic. In *Formal ontology in information systems* (pp. 19–36).

Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2004). *Ontological engineering: With examples from the areas of knowledge management, e-commerce and the semantic web*. Springer.

Hull, C. L. (1920). Quantitative aspects of evolution of concepts: An experimental study. *Psychological Monographs*, *28*(1), i–86.

Imai, M., Saalbach, H., & Stern, E. (2010). Are Chinese and German children taxonomic, thematic, or shape biased? Influence of classifiers and cultural contexts. *Frontier in Psychology*, *1*(194).

Laird, J. E. (2012). *The soar cognitive architecture*. MIT Press.

Margolis, E., & Laurence, S. (Eds.). (2015). *The conceptual mind: New directions in the study of concepts*. MIT Press.

Mitchell, T. M. (1997). *Machine learning*. McGraw Hill.

Murphy, G. L. (2002). *The big book of concepts*. MIT Press.

Murphy, G. L. (2016). Is there an exemplar theory of concepts? *Psychonomic Bulletin & Review*, *23*(4), 1035–1042.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*(3), 289–316.

Potyka, N., & Thimm, M. (2017). Inconsistency-tolerant reasoning over linear probabilistic knowledge bases. *International Journal of Approximate Reasoning*, *88*, 209–236.

Rips, L. J., Bloomfield, A., & Asmuth, J. (2008). From numerical concepts to concepts of number. *Behavioral and Brain Sciences*, *31*(6), 623–642.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT Press.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439.

Russell, S. J., & Norvig, P. (2002). *Artificial intelligence: A modern approach* (International Edition).

Schorlemmer, M., Smaill, A., Kühnberger, K.-U., Kutz, O., Colton, S., Cambouropoulos, E., & Pease, A. (2014). COINVENT: Towards a computational concept invention theory. In *Proceeding of the ICCC*.

Spranger, M. (2012). The co-evolution of basic spatial terms and categories. In L. Steels (Ed.), *Advances in interaction studies experiments in cultural language evolution* (Vol. 3, pp. 111–141). Amsterdam: John Benjamins.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2013). *Intriguing properties of neural networks*.

Generalizing Psychological Similarity Spaces to Unseen Stimuli

Combining Multidimensional Scaling with Artificial Neural Networks

Lucas Bechberger and **Kai-Uwe Kühnberger**

1 Introduction

In this chapter, we propose a combination of psychologically derived similarity ratings with modern machine learning techniques in the context of cognitive artificial intelligence. More specifically, we extract a spatial representation of conceptual similarity from psychological data and learn a mapping from visual input onto this spatial representation.

We base our work on the cognitive framework of conceptual spaces (Gärdenfors 2000), which proposes a geometric representation of conceptual structures: Instances are represented as points and concepts are represented as regions in psychological similarity spaces. Based on this representation, one can explain a range of cognitive phenomena from one-shot learning to concept combination. Conceptual spaces can be interpreted as a spatial variant of the influential prototype theory of concepts (Rosch et al. 1976) by identifying the prototype of a given category with the centroid of the respective convex region. Moreover, conceptual spaces can be related to the feature spaces typically used in machine learning (Mitchell 1997), where individual observations are also represented as sets of feature values and where the task is to identify regions which correspond to pre-defined categories.

As Gärdenfors (2018) has argued, the framework of conceptual spaces splits the overall problem of concept learning into two sub-problems: On the one hand, the space itself with its distance relation and its underlying dimensions needs to be

L. Bechberger (✉) · K.-U. Kühnberger
Institute of Cognitive Science, Osnabrück University, Osnabrück, Germany
e-mail: lucas.bechberger@uni-osnabrueck.de

K.-U. Kühnberger
e-mail: kai-uwe.kuehnberger@uni-osnabrueck.de

learned. On the other hand, one needs to identify meaningful regions within this similarity space. The latter problem can be easily solved by simple learning mechanisms such as taking the centroid of a given set of category members (Gärdenfors 2000). The problem of obtaining the similarity spaces themselves is however much harder. While in humans, the dimensions of these spaces may be partially innate or learned based on perceptual invariants (Gärdenfors 2018), it is difficult to mimic such processes in artificial systems.

When using conceptual spaces as a modeling tool, one can distinguish three ways of obtaining the underlying dimensions: If the domain of interest is well understood, one can manually define the dimensions and thus the overall similarity space. This can for instance be done for the domain of colors, for which a variety of similarity spaces exists. A second approach is based on machine learning algorithms for dimensionality reduction. For instance, unsupervised artificial neural networks (ANNs) such as autoencoders or self-organizing maps can be used to find a compressed representation for a given set of input stimuli. This task is typically solved by optimizing a mathematical error function which may be not satisfactory from a psychological point of view.

A third way of obtaining the dimensions of a conceptual space is based on dissimilarity ratings obtained from human subjects. One first elicits dissimilarity ratings for pairs of stimuli in a psychological study. The technique of "multidimensional scaling" (MDS) takes as an input these pair-wise dissimilarities as well as the desired number t of dimensions. It then represents each stimulus as a point in a t-dimensional space in such a way that the distances between points in this space reflect the dissimilarities of their corresponding stimuli. *Nonmetric* MDS assumes that the dissimilarities are only ordinally scaled and limits itself to representing the ordering of distances correctly. *Metric* MDS on the other hand assumes an interval or ratio scale and also tries to represent the numerical values of the dissimilarities as closely as possible. We introduce multidimensional scaling in more detail in Sect. 2. Moreover, we present a study investigating the differences between similarity spaces produced by metric and nonmetric MDS in Sect. 3.

One limitation of the MDS approach is that it is unable to generalize to unseen inputs: If a new stimulus arrives, it is impossible to directly map it onto a point in the similarity space without eliciting dissimilarities to already known stimuli. In Sect. 4, we propose to use ANNs in order to learn a mapping from raw stimuli to similarity spaces obtained via MDS. This hybrid approach combines the psychological grounding of MDS with the generalization capability of ANNs.

In order to support our proposal, we present the results of a first feasibility study in Sect. 5: Here, we use the activations of a pre-trained convolutional network as features for a simple regression into the similarity spaces from Sect. 3.

Finally, Sect. 6 summarizes the results obtained in this paper and gives an outlook on future work. Code for reproducing both of our studies can be found online at https://github.com/lbechberger/LearningPsychologicalSpaces/ (Bechberger 2020).

Our overall contribution can be seen as providing artificial systems with a way to map raw perceptions onto psychological similarity spaces. These similarity spaces can then be used in order to learn conceptual regions and to reason with them. Our research has strong relations to two other chapters in this edited volume.

The conceptual spaces framework itself can be considered as a specific instance of the approach labeled as "cognitive and distributional semantics" in the contribution by Färber, Svetashova, and Harth (Chap. 3). Our hybrid proposal from Sect. 4 exemplifies the procedure of obtaining such a cognitive representation which is both psychologically grounded and applicable to novel stimuli. Especially the latter property of our hybrid proposal is crucial for applications in technical systems such as the Internet of Things (IoT) considered by Färber, Svetashova, and Harth.

Also the attribute spaces used by Gust and Umbach (Chap. 4) are closely related to the similarity spaces considered in our contribution. While our work focuses on grounding such a similarity space in perception, Gust and Umbach analyze how natural language similarity expressions can be linked to spatial models. The contribution by Gust and Umbach can thus be seen as a complement to our work, considering a higher level of abstraction.

2 Multidimensional Scaling

In this section, we provide a brief introduction to multidimensional scaling. We first give an overview of the elicitation methods for similarity ratings in Sect. 2.1, before explaining the basics of MDS algorithms in Sect. 2.2. The interested reader is referred to Borg and Groenen (2005) for a more detailed introduction to MDS.

2.1 Obtaining Dissimilarity Ratings

In order to collect similarity ratings from human participants, several different techniques can be used (Goldstone 1994; Hout et al. 2013; Wickelmaier 2003). They are typically grouped into *direct* and *indirect* methods: In direct methods, participants are fully aware that they rate, sort, or classify different stimuli according to their pairwise dissimilarities. Indirect methods on the other hand are based on secondary empirical measurements such as confusion probabilities or reaction times.

One of the classical direct techniques is based on explicit ratings for pairwise comparisons. In this approach, all possible pairs from a set of stimuli are presented to participants (one pair at a time), and participants rate the dissimilarity of each pair on a continuous or categorical scale. Another direct technique is based on sorting tasks. For instance, participants might be asked to group a given set of stimuli into piles of similar items. In this case, similarity is binary—either two items are sorted into the same pile or not.

Perceptual confusion tasks can be used as an indirect technique for obtaining similarity ratings. For example, participants can be asked to report as fast as possible whether two displayed items are the same or different. In this case, confusion probabilities and reaction times are measured in order to infer the underlying similarity relation.

Goldstone ([1994](#)) has argued that the classical approaches for collecting similarity data are limited in various ways. Their biggest shortcoming is that explicitly testing all $\frac{N\cdot(N-1)}{2}$ stimulus pairs is quite time-consuming. An increasing number of stimuli therefore leads to very long experimental sessions which might cause fatigue effects. Moreover, in the course of such long sessions, participants might switch to a different rating strategy after some time, making the collected data less homogeneous.

In order to make the data collection process more time-efficient, Goldstone ([1994](#)) has proposed the "*Spatial Arrangement Method*" (SpAM). In this collection technique, multiple visual stimuli are simultaneously displayed on a computer screen. In the beginning, the arrangement of these stimuli is randomized. Participants are then asked to arrange them via drag and drop in such a way that the distances between the stimuli are proportional to their dissimilarities. Once participants are satisfied with their solution, they can store the arrangement. The dissimilarity of two stimuli is then recorded as their Euclidean distance in pixels. As N items can be displayed at once, each single modification by the user updates N distance values at the same time which makes this procedure quite efficient. Moreover, SpAM quite naturally incorporates geometric constraints: If A and B are placed close together and C is placed far away from A, then it cannot be very close to B.

As the dissimilarity information is recorded in the form of Euclidean distances, one might assume that the dissimilarity ratings obtained through SpAM are ratio scaled. This view is for instance held by Hout et al. ([2014](#)). However, as participants are likely to make only a rough arrangement of the stimuli, this assumption might be too strong in practice. One can argue that it is therefore safer to only assume an ordinal scale. As far as we know, there have been no explicit investigations on this issue. We will provide an analysis of this topic in Sect. [3](#).

2.2 The Algorithms

In this chapter, we follow the mathematical notation by Kruskal ([1964a](#)), who gave the first thorough mathematical treatment of (nonmetric) multidimensional scaling.

One can typically distinguish two types of MDS algorithms (Wickelmaier [2003](#)), namely metric and nonmetric MDS. Metric MDS assumes that the dissimilarities are interval or ratio scaled, while nonmetric MDS only assumes an ordinal scale.

Both variants of MDS can be formulated as an optimization problem involving the pairwise dissimilarities δ_{ij} between stimuli and the Euclidean distances d_{ij} of their corresponding points in the t-dimensional similarity space. More specifically, MDS involves minimizing the so-called "stress" which measures to which extent the spatial representation violates the information from the dissimilarity matrix:

$$\text{stress} = \sqrt{\frac{\sum_{i<j}\left(d_{ij}-\hat{d}_{ij}\right)^2}{\sum_{i<j}\left(d_{ij}\right)^2}}$$

The denominator in this equation serves as a normalization factor in order to make stress invariant to the scale of the similarity space.

In metric MDS, we use $\hat{d}_{ij} = a \cdot \delta_{ij} + b$ to compute stress. This means that we look for a configuration of points in the similarity space whose distances are a linear transformation of the dissimilarities.

In nonmetric MDS, on the other hand, the $\hat{d}_{ij}$ are not obtained by a *linear* but by a *monotone* transformation of the dissimilarities: Let us order the dissimilarities of the stimuli ascendingly: $\delta_{i_1 j_1} < \delta_{i_2 j_2} < \delta_{i_3 j_3} < \dots$. The $\hat{d}_{ij}$ are then obtained by defining an analogous ascending order, where the difference between the disparities $\hat{d}_{ij}$ and the distances d_{ij} is as small as possible: $\hat{d}_{i_1 j_1} < \hat{d}_{i_2 j_2} < \hat{d}_{i_3 j_3} < \dots$. Nonmetric MDS therefore only tries to reflect the *ordering* of the dissimilarities in the distances while metric MDS also tries to take into account their differences and ratios.

There are different approaches towards optimizing the stress function, resulting in different MDS algorithms. Kruskal's original nonmetric MDS algorithm (Kruskal 1964b) is based on gradient descent: In an iterative procedure, the derivative of the stress function with respect to the coordinates of the individual points is computed and then used to make a small adjustment to these coordinates. Once the derivative approaches zero, a minimum of the stress function has been found.

A more recent MDS algorithm by de Leeuw (1977) is called SMACOF (an acronym of "**S**caling by **Ma**jorizing a **Co**mplicated **F**unction"). De Leeuw pointed out that Kruskal's gradient descent method has two major shortcomings: Firstly, if the points for two stimuli coincide (i.e., $x_i = x_j$), then the distance function of these two points is not differentiable. Secondly, Kruskal was not able to give a proof of convergence for his algorithm. In order to overcome these limitations, De Leeuw showed that minimizing the stress function is equivalent to maximizing another function λ which depends on the distances and dissimilarities. This function can be easily maximized by using iterative function majorization. Moreover, one can prove that this iterative procedure converges. SMACOF is computationally efficient and guarantees a monotone convergence of stress (Borg and Groenen 2005, Chap. 8).

Picking the right number of dimensions t for the similarity space is not trivial. Kruskal (1964a) proposes two approaches to address this problem.

On the one hand, one can create a so-called "Scree" plot that shows the final stress value for different values of t. If one can identify an "elbow" in this diagram (i.e., a point after which the stress decreases much slower than before), this can point towards a useful value of t.

On the other hand, one can take a look at the interpretability of the generated configurations. If the optimal configuration in a t-dimensional space has a sufficient degree of interpretability and if the optimal configuration in a $t + 1$ dimensional space does not add more structure, then a t-dimensional space might be sufficient.

Fig. 1 Eight example stimuli from the NOUN data set (Horst and Hout 2016)

3 Extracting Similarity Spaces from the NOUN Data Set

It is debatable whether metric or nonmetric MDS should be used with data collected through SpAM. Nonmetric MDS makes less assumptions about the underlying measurement scale and therefore seems to be the "safer" choice. If the dissimilarities are however ratio scaled, then metric MDS might be able to harness these pieces of information from the distance matrix as additional constraints. This might then result in a semantic space of higher quality.

In our study, we compare metric to nonmetric MDS on a data set obtained through SpAM. If the dissimilarities obtained through SpAM are not ratio scaled, then the main assumption of metric MDS is violated. We would then expect that nonmetric MDS yields better solutions than metric MDS. If the dissimilarities obtained through SpAM are however ratio scaled and if the differences and ratios of dissimilarities do contain considerable amounts of additional information, then metric MDS should have a clear advantage over nonmetric MDS.

For our study, we used existing dissimilarity ratings reported for the Novel Object and Unusual Name (NOUN) data set (Horst and Hout 2016), a set of 64 images of three-dimensional objects that are designed to be novel but also look naturalistic. Figure 1 shows some example stimuli from this data set.

3.1 Evaluation Metrics

We used the `stress0` function from R's `smacof` package to compute both metric and nonmetric stress. We expect stress to decrease as the number of dimensions increases. If the data obtained through SpAM is ratio scaled, then we would expect that metric MDS achieves better values on metric stress (and potentially on nonmetric stress as well) than nonmetric MDS. If the SpAM dissimilarities are not ratio scaled, then metric MDS should not have any advantage over nonmetric MDS.

Another possible way of judging the quality of an MDS solution is to look for interpretable directions in the resulting space. However, Horst and Hout (2016) have argued that for the novel stimuli in their data set there are no obvious directions that one would expect. Without a list of candidate directions, an efficient and objective evaluation based on interpretable directions is however hard to achieve. We therefore did not pursue this way of evaluating similarity spaces.

As an additional way of evaluation, we measured the correlation between the distances in the MDS space and the dissimilarity scores from the psychological study.

Pearson's *r* (Pearson 1895) measures the linear correlation of two random variables by dividing their covariance by the product of their individual variances. Given two vectors *x* and *y* (each containing *N* samples from the random variables *X* and *Y*, respectively), Pearson's *r* can be estimated as follows, where $\bar{x}$ and $\bar{y}$ are the average values of the two vectors:

$$r_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

Spearman's ρ (Spearman 1904) generalizes Pearson's *r* by allowing also for nonlinear monotone relationships between the two variables. It can be computed by replacing each observation x_i and y_i with its corresponding rank, i.e., its index in a sorted list, and by then computing Pearson's *r* on these ranks. By replacing the actual values with their ranks, the numeric distances between the sample values lose their importance—only the correct ordering of the samples remains important. Like Pearson's *r*, Spearman's ρ is confined to the interval $[-1, 1]$ with positive values indicating a monotonically increasing relationship.

Both MDS variants can be expected to find a configuration such that there is a monotone relationship between the distances in the similarity space and the original dissimilarity matrix. That is, smaller dissimilarities correspond to smaller distances and larger dissimilarities correspond to larger distances. For Spearman's ρ, we therefore do not expect any notable differences between metric and nonmetric MDS. For metric MDS, we also expect a *linear* relationship between dissimilarities and distances. Therefore, if the dissimilarities obtained by SpAM are ratio scaled, then metric MDS should give better results with respect to Pearson's *r* than nonmetric MDS.

A final way for evaluating the similarity spaces obtained by MDS is visual inspection: If a visualization of a given similarity space shows meaningful structures and clusters, this indicates a high quality of the semantic space. We limit our visual inspection to two-dimensional spaces.

3.2 Methods

In order to investigate the differences between metric and nonmetric MDS in the context of SpAM, we used the SMACOF algorithm in its original implementation in R's `smacof` library.[1] SMACOF can be used in both a *metric* and a *nonmetric* variant. The underlying algorithm stays the same, only the definition of stress and

---

[1]See https://cran.r-project.org/web/packages/smacof/smacof.pdf.

thus the optimization target differs. Both variants were explored in our study. We used 256 random starts with the maximum number of iterations per random start set to 1000. The overall best result over these 256 random starts was kept as final result.

For each of the two MDS variants, we constructed MDS spaces of different dimensionality (ranging from one to ten dimensions). For each of these resulting similarity spaces, we computed both its metric and its nonmetric stress.

In order to analyze how much information about the dissimilarities can be readily extracted from the images of the stimuli, we also introduced two baselines.

For our first baseline, we used the similarity of downscaled images: For each original image (with both a width and height of 300 pixels), we created lower-resolution variants by aggregating all the pixels in a $k \times k$ block into a single pixel (with $k \in [2, 300]$). We compared different aggregation functions, namely, minimum, mean, median, and maximum. The pixels of the resulting downscaled image were then interpreted as a point in a $\lceil \frac{300}{k} \rceil \times \lceil \frac{300}{k} \rceil$ dimensional space.

For our second baseline, we extracted the activation vectors from the second-to-last layer of the pre-trained Inception-v3 network (Szegedy et al. 2016) for each of the images from the NOUN data set. Each stimulus was thus represented by its corresponding activation pattern. While the downscaled images represent surface level information, the activation patterns of the neural network can be seen as more abstract representation of the image.

For each of the three representation variants (downscaled images, ANN activations, and points in an MDS-based similarity space), we computed three types of distances between all pairs of stimuli: The Euclidean distance d_E, the Manhattan distance d_M, and the negated inner product d_{IP}. We only report results for the best choice of the distance function. For each distance function, we used two variants: One where all dimensions are weighted equally and another one where optimal weights for the individual dimensions were estimated based on a non-negative least squares regression in a five-fold cross validation (cf. Peterson et al. (2018) who followed a similar procedure). For each of the resulting distance matrices, we compute the two correlation coefficients with respect to the target dissimilarity ratings. We consider only matrix entries above the diagonal because the matrices are symmetric and all entries on the diagonal are guaranteed to be zero. Our overall workflow is illustrated in Fig. 2.

3.3 Results

Figure 3a shows the Scree plots of the two MDS variants for both metric and nonmetric stress. As one would expect, stress decreases with an increasing number of dimensions: More dimensions help to represent the dissimilarity ratings more accurately. Metric and nonmetric SMACOF yield almost identical performance with respect to both metric and nonmetric stress. This suggests that interpreting the SpAM dissimilarity ratings as ratio scaled is neither helpful nor harmful.
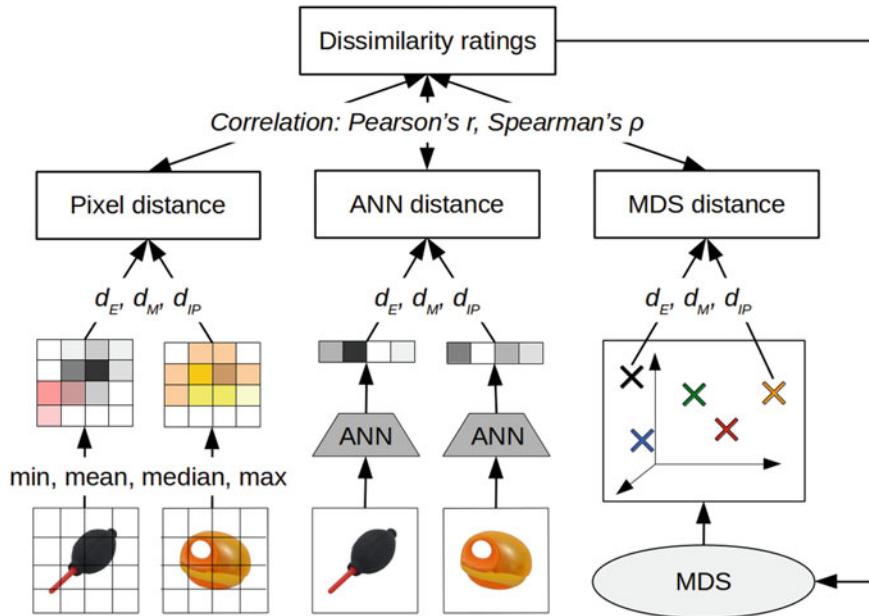
Fig. 2 Illustration of our analysis setup. We measure the correlation between the dissimilarity ratings and distances from three different sources, namely the pixels of downscaled images (left), activations of an artificial neural network (middle), and similarity spaces obtained by MDS (right)

Figure 3b shows some line diagrams illustrating the results of the correlation analysis for the MDS-based similarity spaces. For both the pixel baseline and the ANN baseline, the usage of optimized weights considerably improved performance. As we can see, both of these baselines yield considerably higher correlations than one would expect for randomly generated configurations of points. Moreover, the ANN baseline outperforms the pixel baseline with respect to both evaluation metrics, indicating that raw pixel information is less useful in our scenario than the more high-level features extracted by the ANN. For the pixel baseline, we observed that the minimum aggregator yielded the best results.

We also observe in Fig. 3b that the MDS solutions provide us with a better reflection of the dissimilarity ratings than both pixel-based and ANN-based distances if the similarity space has at least two dimensions. This is not surprising since the MDS solutions are directly based on the dissimilarity ratings, whereas both baselines do not have access to the dissimilarity information. It therefore seems like our naive image-based ways of defining dissimilarities are not sufficient.

With respect to the different MDS variants, also the correlation analysis confirms our observations from the Scree plots: Metric and nonmetric SMACOF are almost indistinguishable with nonmetric SMACOF yielding slightly higher correlation values. This supports the view that the assumption of ratio scaled dissimilarity ratings is not beneficial, but also not very harmful on out data set. Moreover, we find the
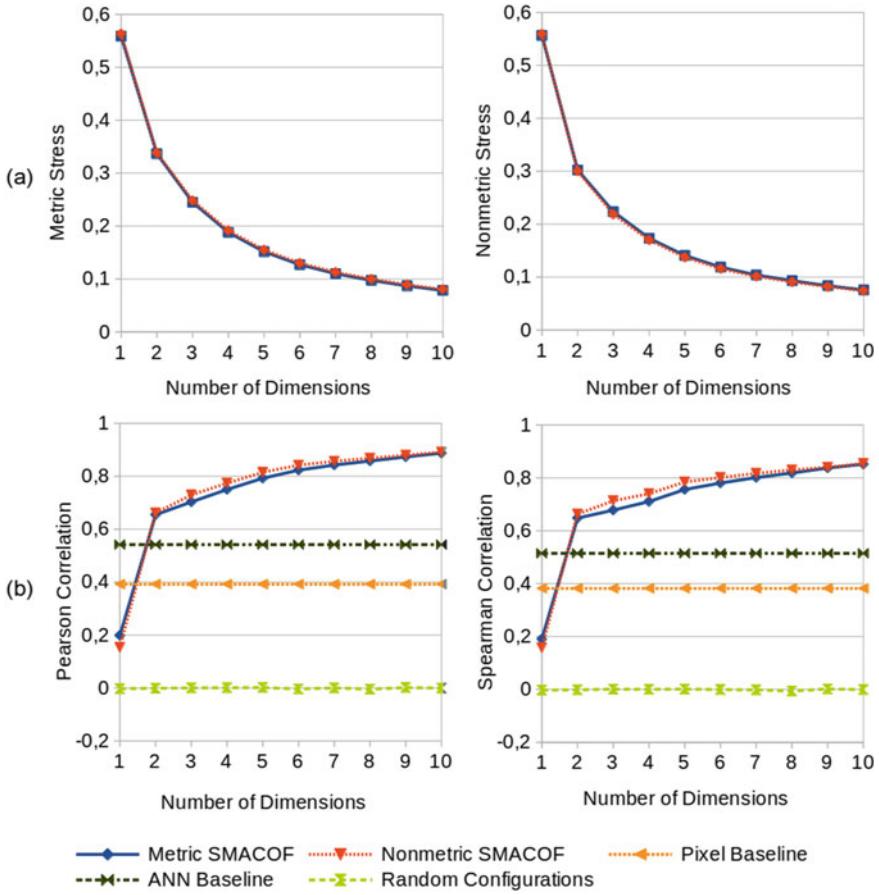
Fig. 3 a Scree plots for both metric and nonmetric stress. **b** Correlation evaluation for the different MDS solutions and the two baselines

tendency of improved performance with an increasing number of dimensions. This again illustrates that MDS is able to fit more information into the space if this space has a larger dimensionality.

Finally, let us look at the two-dimensional spaces generated by the two MDS variants in order to get an intuitive feeling for their semantic structure. Figure 4 shows these spaces along with the local neighborhood of three selected items. These neighborhoods illustrate that in both spaces stimuli are grouped in a meaningful way. From our visual inspection, it seems that both MDS variants result in comparable semantic spaces with a similar structure.

Overall, we did not find any systematic difference between metric and nonmetric MDS on the given data set. It thus seems that the metric assumption is neither beneficial nor harmful when trying to extract a similarity space. On the one hand, we
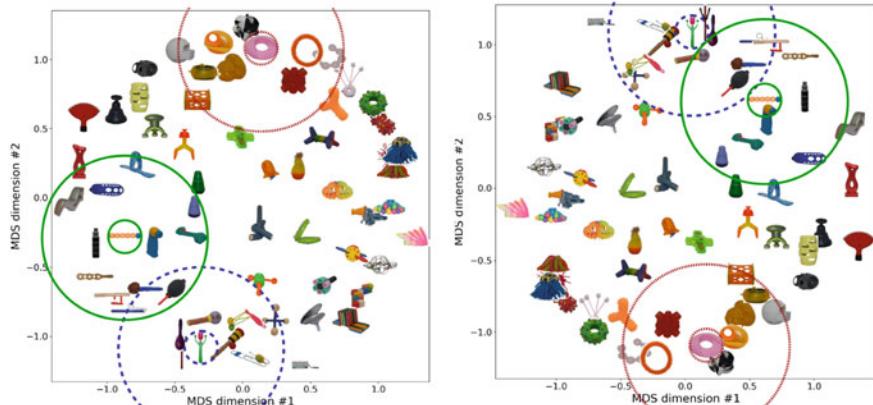
Fig. 4 Illustration of the two-dimensional spaces obtained by metric SMACOF (left) and nonmetric SMACOF (right)

cannot conclude that the dissimilarities obtained through SpAM are *not* ratio scaled. On the other hand, the additional information conveyed by differences and ratios of dissimilarities does not seem to improve the overall results. We therefore advocate the usage of nonmetric MDS due to the smaller amount of assumptions made about the dissimilarity ratings.

4 A Hybrid Approach

Multidimensional scaling (MDS) is directly based on human similarity ratings and leads therefore to conceptual spaces which can be considered psychologically valid. The prohibitively large effort required to elicit such similarity ratings on a large scale however confines this approach to a small set of fixed stimuli. In Sect. 4.1, we propose to use machine learning methods in order to generalize the similarity spaces obtained by MDS to unseen stimuli. More specifically, we propose to use MDS on human similarity ratings to "initialize" the similarity space and artificial neural networks (ANNs) to learn a mapping from stimuli into this similarity space. We afterwards relate our proposal to two other recent studies in this area in Sect. 4.2.

4.1 Our Proposal

In order to obtain a solution having both the psychological validity of MDS spaces and the possibility to generalize to unseen inputs as typically observed for neural networks, we propose the following hybrid approach, which is illustrated in Fig. 5.
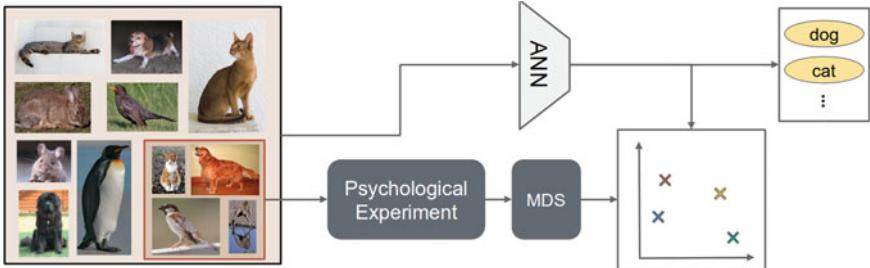
Fig. 5 Illustration of the proposed hybrid procedure: a subset of data is used to construct a conceptual space via MDS. A neural network is then trained to map images into this similarity space, aided by a secondary task (e.g., classification)

After having determined the domain of interest (e.g., the domain of animals), one first needs to acquire a data set of stimuli from this domain. This data set should cover a wide variety of stimuli and it should be large enough for applying machine learning algorithms. Using the whole data set with potentially thousands of stimuli in a psychological experiment is however unfeasible in practice. Therefore, a relatively small, but still sufficiently representative subset of these stimuli needs to be selected for the elicitation of human dissimilarity ratings. This subset of stimuli is then used in a psychological experiment where dissimilarity judgments by humans are obtained, using one of the techniques described in Sect. 2.1.

In the next step, one can apply MDS to these dissimilarity ratings in order to extract a spatial representation of the underlying domain. As stated in Sect. 2.2, one needs to manually select the desired number of dimensions—either based on prior knowledge or by manually optimizing the trade-off between high representational accuracy and a low number of dimensions. The resulting similarity space should ideally be analyzed for meaningful structures and a high correlation of inter-point distances to the original dissimilarity ratings.

Once this mapping from stimuli (e.g., images of animals) to points in a similarity space has been established, we can use it in order to derive a ground truth for a machine learning problem: We can simply treat the stimulus-point mappings as labeled training instances where the stimulus is identified with the input vector and the point in the similarity space is used as its label. We can therefore set up a regression task from the stimulus space to the similarity space.

Artificial neural networks (ANNs) have been shown to be powerful regressors that are capable of discovering highly non-linear relationships between raw low-level stimuli (such as images) and desired output variables. They are therefore a natural choice for this task. ANNs are however a very data-hungry machine learning method — they need large amounts of training examples and many training iterations in order to achieve good performance. On the other hand, the available number of stimulus-point pairs in our proposed procedure is quite low for a machine learning problem — as argued before, we can only look at a small number of stimuli in a psychological experiment.

We propose to resolve this dilemma not only through data augmentation, but also by introducing an additional training objective (e.g., correctly classifying the given images into their respective classes such as CAT and DOG). This additional training objective can also be optimized on all the remaining stimuli from the data set that have not been used in the psychological experiment. Using a secondary task with additional training data constrains the network's weights and can be seen as a form of regularization: These additional constraints are expected to counteract overfitting tendencies, i.e., tendencies to memorize all given mapping examples without being able to generalize.

Figure 5 illustrates the secondary task of predicting the correct classes. This approach is only applicable if the data set contains class labels. If the network is forced to learn a classification task, then it will likely develop an internal representation where all members of the same class are represented in a similar way. The network then "only" needs to learn a mapping from this internal representation (which presumably already encodes at least some aspects of a similarity relation between stimuli) into the target similarity space.

Another secondary task consists in reconstructing the original images from a low-dimensional internal representation, using the structure of an autoencoder. As the computation of the reconstruction error does not require class labels, this is applicable also to unlabeled data sets, which are in general larger and easier to obtain than labeled data sets. The network needs to accurately reconstruct the given stimuli while using only information from a small bottleneck layer. The small size of the bottleneck layer creates an incentive to encode similar input stimuli in similar ways such that the corresponding reconstructions are also similar to each other. Again, this similarity relation learned from the overall data set might be useful for learning the mapping into the similarity space. The autoencoder structure has the additional advantage that one can use the decoder network to generate an image based on a point in the conceptual space. This can be a useful tool for visualization and further analysis.

One should be aware that there is a difference between perceptual and conceptual similarity: Perceptual similarity focuses on the similarity of the raw stimuli, e.g., with respect to their shape, size, and color. Conceptual similarity on the other hand takes place on a more abstract level and involves conceptual information such as the typical usage of an object or typical locations where a given object might be found. For instance, a violin and a piano are perceptually not very similar as they have different sizes and shapes. Conceptually, they might be however quite similar as they are both musical instruments that can be found in an orchestra.

While class labels can be assigned on both the perceptual (ROUND vs. ELON-GATED) and the conceptual level (MUSICAL INSTRUMENT vs. FRUIT), the reconstruction objective always operates on the perceptual level. If the similarity data collected in the psychological experiment is of perceptual nature, then both secondary tasks seem promising. If we however target conceptual similarity, then the classification objective seems to be the preferable choice.

4.2 Related Work

Peterson et al. (2017, 2018) have investigated whether the activation vectors of a neural network can be used to predict human similarity ratings. They argue that this can enable researchers to validate psychological theories on large data sets of real world images.

In their study, they used six data sets containing 120 images (each 300 by 300 pixels) of one visual domain (namely, animals, automobiles, fruits, furniture, vegetables, and "various"). Peterson et al. conducted a psychological study which elicited pairwise similarity ratings for all pairs of images using a Likert scale. When applying multidimensional scaling to the resulting dissimilarity matrix, they were able to identify clear clusters in the resulting space (e.g., all birds being located in a similar region of the animal space). Moreover, when applying a hierarchical clustering algorithm on the collected similarity data, a meaningful dendrogram emerged.

In order to extract similarity ratings from five different neural networks, they computed for each image the activation in the second-to-last layer of the network. Then for each pair of images, they defined their similarity as the inner product ($u^T v = \sum_{i=1}^{n} u_i v_i$) of these activation vectors. When applying MDS to the resulting dissimilarity matrix, no meaningful clusters were observed. Also a hierarchical clustering did not result in a meaningful dendrogram. When considering the correlation between the dissimilarity ratings obtained from the neural networks and the human dissimilarity matrix, they were able to achieve values of R^2 between 0.19 and 0.58 (depending on the visual domain).

Peterson et al. found that their results considerably improved when using a weighted version of the inner product ($\sum_{i=1}^{n} w_i u_i v_i$): Both the similarity space obtained by MDS and the dendrogram obtained by hierarchical clustering became more human-like. Moreover, the correlation between the predicted similarities and the human similarity ratings increased to values of R^2 between 0.35 and 0.74.

While the approach by Peterson et al. illustrates that there is a connection between the features learned by neural networks and human similarity ratings, it differs from our proposed approach in one important aspect: Their primary goal is to find a way to predict the similarity ratings directly. Our research on the other hand is focused on predicting points in the underlying similarity space.

Sanders and Nosofsky (2018) have used a data set containing 360 pictures of rocks along with an eight-dimensional similarity space for a study which is quite similar in spirit to what we will present in Sect. 5. Their goal was to train an ensemble of convolutional neural networks for predicting the correct coordinates in the similarity space for each rock image from the data set. As the data set is considerably too small for training an ANN from scratch, they used a pre-trained network as a starting point. They removed the topmost layers and replaced them by untrained, fully connected layers with an output of eight linear units, one per dimension of the similarity space. In order to increase the size of their data set, they applied data augmentation methods by flipping, rotating, cropping, stretching, and shrinking the original images.

Their results on the test set showed a value of R^2 of 0.808, which means that over 80% of the variance was accounted for by the neural network. Moreover, an exemplar model on the space learned by the convolutional neural network was able to explain 98.9% of the variance seen in human categorization performance.

The work by Sanders and Nosofsky is quite similar in spirit to our own approach: Like we, they train a neural network to learn the mapping between images and a similarity space extracted from human similarity ratings. They do so by resorting to a pre-trained neural network and by using data augmentation techniques. While they use a data set of 360 images, we are limited to an even smaller data set containing only 64 images. This makes the machine learning problem even more challenging. Moreover, the data set used by Sanders and Nosofky is based on real objects, whereas our study investigates a data set of novel and unknown objects. Finally, while they confine themselves to a single target similarity space for their regression task, we investigate the influence of the target space on the overall results.

5 Machine Learning Experiments

In order to validate whether our proposed approach is worth pursuing, we conducted a feasibility study based on the similarity spaces obtained for the NOUN data set in Sect. 3. Instead of training a neural network from scratch, we limit ourselves to a simple regression on top of a pre-trained image classification network. With the three experiments in our study, we address the following three research questions, respectively:

1. Can we learn a useful mapping from colored images into a low-dimensional psychological similarity space from a small data set of novel objects for which no background knowledge is available?
   Our prediction: The learned mapping is able to clearly beat a simple baseline. However, it does not reach the level of generalization observed in the study of Sanders and Nosofsky (2018) *due to the smaller amount of data available.*
2. How does the MDS algorithm being used to construct the target similarity space influence the results?
   Our prediction: There is are no considerable differences between metric and nonmetric MDS.
3. How does the size of the target similarity space (i.e., the number of dimensions) influence the machine learning results?
   Our prediction: Very small target spaces are not able to reflect the similarity ratings very well and do not contain much meaningful structure. Very large target spaces on the other hand increase the number of parameters in the model which makes overfitting more likely. By this reasoning, medium-sized target spaces should provide a good trade-off and therefore the best regression performance.

5.1 Methods

Please recall from Sect. 3 that the NOUN data base contains only 64 images with an image size of 300 by 300 pixels. As this number of training examples is too low for applying machine learning techniques, we augmented the data set by applying random crops, a Gaussian blur, additive Gaussian noise, affine transformations (i.e., rotations, shears, translations, and scaling), and by manipulating the image's contrast and brightness. These augmentation steps were executed in random order and with randomized parameter settings. For each of the original 64 images, we created 1,000 augmented versions, resulting in a data set of 64,000 images in total. We assigned the target coordinates of the original image to each of the 1,000 augmented versions.

For our regression experiments, we used two different types of feature spaces: The pixels of downscaled images and high-level activation vectors of a pre-trained neural network.

For the ANN-based features, we used the Inception-v3 network (Szegedy et al. 2016). For each of the augmented images, we used the activations of the second-to-last layer as a 2048-dimensional feature vector. Instead of training both the mapping and the classification task simultaneously (as discussed in Sect. 4), we use an already pre-trained network and augment it by an additional output layer.

As a comparison to the ANN-based features, we used an approach similar to the pixel baseline from Sect. 3.2: We downscaled each of the augmented images by dividing it into equal-sized blocks and by computing the minimum (which has shown the best correlation to the dissimilarity ratings in Sect. 3.3) across all values in each of these blocks as one entry of the feature vector. We used block sizes of 12 and 24, resulting in feature vectors of size 1875 and 507, respectively (based on three color channels for downscaled images of size 25×25 and 13×13, respectively). By using these two pixel-based feature spaces, we can analyze differences between low-dimensional and high-dimensional feature spaces. As the high-dimensional feature space is in the same order of magnitude as the ANN-based feature space, we can also make a meaningful comparison between pixel-based features and ANN-based features.

We compare our regression results to the zero baseline which always predicts the origin of the coordinate system. In preliminary experiments, it has shown to be superior to any other simple baselines (such as e.g., drawing from a normal distribution estimated from the training targets). We do not expect this baseline to perform well in our experiments, but it defines a lower performance bound for the regressors.

In our experiments, we limit ourselves to two simple off-the-shelf regressors, namely a linear regression and a lasso regression. Let N be the number of data points, t be the number of target dimensions, $y_d^{(i)}$ the target value of data point i in dimension d, and $f_d^{(i)}$ the prediction of the regressor for data point i in dimension d.

Both of our regressors make use of a simple linear model for each of the dimensions in the target space:

$$f_d = w_0^{(d)} + \sum_{k=1}^{K} w_k^{(d)} x_k$$

Here, K is the number of features and x is the feature vector. In a linear least-squares regression, the weights $w_k^{(d)}$ of this model are estimated by minimizing the mean squared error between the model's predictions and the actual ground truth value:

$$MSE_d = \frac{1}{N} \sum_{i=1}^{N} \left(y_d^{(i)} - f_d^{(i)} \right)^2$$

As the number of features is quite high, even a linear regression needs to estimate a large number of weights. In order to prevent overfitting, we also consider a lasso regression which additionally incorporates the L_1 norm of the weight matrix as regularization term. It minimizes the following objective:

$$\frac{1}{N} \sum_{i=1}^{N} \left(y_d^{(i)} - f_d^{(i)} \right)^2 + \beta \cdot \frac{1}{K} \cdot \sum_{k=1}^{K} w_k^{(d)}$$

The first part of this objective corresponds to the mean squared error of the linear model's predictions, while the second part corresponds to the overall size of the weights. If the constant β is tuned correctly, this can prevent overfitting and thus improve performance on the test set. In our experiments, we investigated the following values:

$$\beta \in \{0.0, 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1.0, 2.0, 5.0, 10.0\}$$

Please note that $\beta = 0$ corresponds to an ordinary linear least-squares regression.

With our experiments, we would also like to investigate whether learning a mapping into a psychological similarity space is easier than learning a mapping into an arbitrary space of the same dimensionality. In addition to the real regression targets (which are the coordinates from the similarity space obtained by MDS), we created another set of regression targets by randomly shuffling the assignment from images to target points. We ensured that all augmented images created from the same original image were still mapped onto the same target point. With this shuffling procedure, we aimed to destroy any semantic structure inherent in the target space. We expect that the regression works better for the original targets than for the shuffled targets.

In order to evaluate both the regressors and the baseline, we used three different evaluation metrics:

- The **mean squared error (MSE)** sums over the average squared difference between the prediction and the ground truth for each output dimension.

$$MSE = \sum_{d=1}^{t} \frac{1}{N} \cdot \sum_{i=1}^{N} \left(y_d^{(i)} - f_d^{(i)} \right)^2$$

- The **mean euclidean distance (MED)** provides us with a way of quantifying the average distance between the prediction and the target in the similarity space.

$$MED = \frac{1}{N} \cdot \sum_{i=1}^{N} \sqrt{\sum_{d=1}^{t} \left(y_d^{(i)} - f_d^{(i)} \right)^2}$$

- The **coefficient of determination** R^2 can be interpreted as the amount of variance in the targets that is explained by the regressor's predictions.

$$R^2 = \frac{1}{t} \cdot \sum_{d=1}^{t} \left(1 - \frac{S_{residual}^{(d)}}{S_{total}^{(d)}} \right) \text{ with } S_{residual}^{(d)} = \sum_{i=1}^{N} \left(y_d^{(i)} - f_d^{(i)} \right)^2$$

$$\text{and } S_{total}^{(d)} = \sum_{i=1}^{N} \left(y_d^{(i)} - \bar{y} \right)^2$$

We evaluated all regressors using an eight-fold cross validation approach, where each fold contains all the augmented images generated from eight of the original images. In each iteration, one of these folds was used as test set, whereas all other folds were used as training set. We aggregated all predictions over these eight iterations (ending up with exactly one prediction per data point) and computed the evaluation metrics on this set of aggregated predictions.

5.2   Experiment 1: Comparing Feature Spaces and Regressors

In our first experiment, we want to test the following hypotheses:

1. The learned mapping is able to clearly beat the baseline. However, it does not reach the level of generalization observed in the study of Sanders and Nosofsky (2018) due to the smaller amount of data available.
2. A regression from the ANN-based features is more successful than a regression from the pixel-based features.
3. As the similarity spaces created by MDS encode semantic similarity by geometric distance, we expect that learning the correct mapping generalizes better to the test set than learning a shuffled mapping.
4. As the feature vectors are quite large, the linear regression has a large number of weights to optimize, inviting overfitting. Regularization through the L_1 loss included in the lasso regressor can help to reduce overfitting.

Table 1 Performance of the different regressors for different feature spaces and correct versus shuffled targets on the four-dimensional space by Horst and Hout (2016). The best results for each combination of column and regressor are highlighted in boldface

Regressor	Feature space	Targets	Test set performance			Degree of overfitting			β
			MSE	MED	R^2	MSE	MED	R^2	
Baseline	Any	Any	1.0000	0.9962	0.0000	1.0000	1.0000	1.0000	–
Linear	ANN (2048)	Correct	**0.6076**	**0.7498**	**0.3766**	42.7093	6.8876	**2.6171**	–
		Shuffled	1.1440	1.0505	−0.1390	56.5298	8.1034	−7.0475	–
	Pixel (1875)	Correct	1.3172	1.0845	−0.3251	2.6191	1.6310	−1.5199	–
		Shuffled	1.6075	1.2099	−0.5955	2.6629	1.6522	−0.6700	–
	Pixel (507)	Correct	1.2073	1.0428	−0.2120	**2.3360**	**1.5433**	−2.2664	–
		Shuffled	1.5853	1.2072	−0.5727	2.5049	1.6036	−0.6458	–
Lasso	ANN (2048)	Correct	**0.5740**	**0.7264**	**0.4111**	28.3766	5.6239	2.3727	0.005, 0.01
	Pixel (1875)	Correct	0.9183	0.9391	0.0788	**1.1320**	1.1371	2.3313	0.2, 0.5
	Pixel (507)	Correct	0.8946	0.9292	0.1015	1.1677	**1.1251**	**2.2538**	0.05, 0.1

5. For smaller feature vectors, we expect less overfitting tendencies than for larger feature vectors. Therefore, less regularization should be needed to achieve optimal performance.

Here, we limit ourselves to a single target space, namely the four-dimensional similarity space obtained by Horst and Hout (2016) through metric MDS.

Table 1 shows the results obtained in our experiment, grouped by the regression algorithm, feature space, and target mapping used. We have also reported the observed degree of overfitting. It is calculated by dividing training set performance by test set performance. Perfect generalization would result in a degree of overfitting of one, whereas larger values reflect the factor to which the regression is more successful on the training set than on the test set. Let us for now only consider the linear regression.

We first focus on the results obtained on the ANN-based feature set. As we can see, the linear regression is able to beat the baseline when trained on the correct targets. The overall approach therefore seems to be sound. However, we see strong overfitting tendencies, showing that there is still room for improvement. When trained on the shuffled targets, the linear regression completely fails to generalize to the test set. This shows that the correct mapping (having a semantic meaning) is easier to learn than an unstructured mapping. In other words, the semantic structure of the similarity space makes generalization possible.

Let us now consider the pixel-based feature spaces. For both of these spaces, we observe that linear regression performs worse than the baseline. Moreover, we can see that learning the shuffled mapping results in even poorer performance than

learning the correct mapping. Due to the overall poor performance, we do not observe very strong overfitting tendencies. Finally, when comparing the two pixel-based feature spaces, we observe that the linear regression tends to perform better on the low-dimensional feature space than on the high-dimensional one. However, these performance differences are relatively small.

Overall, ANN-based features seem to be much more useful for our mapping task than the simple pixel-based features, confirming our observations from Sect. 3.

In order to further improve our results, we now varied the regularization factor β of the lasso regressor for all feature spaces.

For the ANN-based feature space, we are able to achieve a slight but consistent improvement by introducing a regularization term: Increasing β causes poorer performance on the training set while yielding improvements on the test set. The best results on the test set are achieved for $\beta \in \{0.005, 0.01\}$. If β however becomes too large, then performance on the test set starts to decrease again — for $\beta = 0.05$ we do not see any improvements over the vanilla linear regression any more. For $\beta \geq 5$, the lasso regression collapses and performs worse than the baseline.

Although we are able to improve our performance slightly, the gap between training set performance and test set performance still remains quite high. It seems that the overfitting problem can be somewhat mitigated but not solved on our data set with the introduction of a simple regularization term.

When comparing our best results to the ones obtained by Sanders and Nosofsky (2018) who achieved values of $R^2 \approx 0.8$, we have to recognize that our approach performs considerably worse with $R^2 \approx 0.4$. However, the much smaller number of data points in our experiment makes our learning problem much harder than theirs. Even though we use data augmentation, the small number of different targets might put a hard limit on the quality of the results obtainable in this setting. Moreover, Sanders and Nosofsky retrained the whole neural network in their experiments, whereas we limit ourselves to the features extracted by the pre-trained network. As we are nevertheless able to clearly beat our baselines, we take these results as supporting the general approach.

For the pixel-based feature spaces, we can also observe positive effects of regularization. For the large space, the best results on the test set are achieved for larger values of $\beta \in \{0.2, 0.5\}$. These results are however only slightly better than baseline performance. For the small pixel-based feature space, the optimal value of β lies in $\{0.05, 0.1\}$, leading again to a test set performance slightly superior to the baseline. In case of the small pixel-based feature space, already values of $\beta \geq 1$ lead to a collapse of the model.

Comparing the regularization results on the three feature spaces, we can conclude that regularization is indeed helpful, but only to a small degree. On the ANN-based feature space, we still observe a large amount of overfitting, and performance on the pixel-based feature spaces is still relatively close to the baseline. Looking at the optimal values of β, it seems like the lower-dimensional pixel-based feature space needs less regularization than its higher-dimensional counterpart. Presumably, this is caused by the smaller possibility for overfitting in the lower-dimensional feature space. Even though the larger pixel-based feature space and the ANN-based feature

space have a similar dimensionality, the pixel-based feature space requires a larger degree of regularization for obtaining optimal performance, indicating that it is more prone to overfitting than the ANN-based feature space.

5.3 Experiment 2: Comparing MDS Algorithms

After having analyzed the soundness of our approach in experiment 1, we compare target spaces of the same dimensionality, but obtained with different MDS algorithms. More specifically, we compare the results from experiment 1 to analogous procedures applied to the ANN-based feature space and the four-dimensional similarity spaces created by both metric and nonmetric SMACOF in Sect. 3. Table 2 shows the results of our second experiment.

In a first step, we can compare the different target spaces by taking a look at the behavior of the zero baseline in each of them. As we can see, the values for MSE and R^2 are identical for all of the different spaces. Only for the MED we can observe some slight variations, which can be explained by the slightly different arrangements of points in the different similarity spaces.

As we can see from Table 2, the results for the linear regression on the different target spaces are comparable. This adds further support to our results from Sect. 3:

Table 2 Comparison of the results obtainable on four-dimensional spaces created by different MDS algorithms. Best results in each column are highlighted for each of the regressors

Regressor	Target space	Test set performance			Amount of overfitting			β
		MSE	MED	R^2	MSE	MED	R^2	
Baseline	Horst and Hout	1.0000	0.9962	0.0000	1.0000	1.0000	1.0000	–
	Metric SMACOF	1.0000	0.9981	0.0000	1.0000	1.0000	1.0000	–
	Nonmetric SMACOF	1.0000	**0.9956**	0.0000	1.0000	1.0000	1.0000	–
Linear	Horst and Hout	**0.6076**	0.7498	**0.3766**	42.7093	6.8876	2.6171	–
	Metric SMACOF	0.6172	0.7560	**0.3766**	**42.2885**	6.8583	**2.6162**	–
	Nonmetric SMACOF	0.6086	**0.7461**	0.3706	42.4380	**6.8305**	2.6585	–
Lasso	Horst and Hout	**0.5740**	**0.7264**	**0.4111**	**28.3766**	**5.6239**	**2.3727**	0.005, 0.01
	Metric SMACOF	0.6052	0.7458	0.3880	35.0367	6.2463	2.5326	0.002
	Nonmetric SMACOF	0.5938	0.7316	0.3853	29.1236	5.6497	2.5413	0.005

Also when considering the usage as target space for machine learning, metric MDS does not seem to have any advantage over nonmetric MDS.

For the lasso regressor, we observed similar effects for all of the target spaces: A certain amount of regularization is helpful to improve test set performance, while too much emphasis on the regularization term causes both training and test set performance to collapse. We still observe a large amount of overfitting even after using regularization. Again, the results are comparable across the different target spaces. However, the optimal performance on the space obtained with metric SMACOF is consistently worse than the results obtained on the other two spaces. As the space by Horst and Hout is however also based on metric MDS, we cannot use this observation as an argument for nonmetric MDS.

5.4   Experiment 3: Comparing Target Spaces of Different Size

In our third and final experiment in this study, we vary the number of dimensions in the target space. More specifically, we consider similarity spaces with one to ten dimensions that have been created by nonmetric SMACOF. Again, we only consider the ANN-based feature space.

Table 3 displays the results obtained in our third experiment and Fig. 6 provides a graphical illustration. When looking at the zero baseline, we observe that the mean Euclidean distance tends to grow with an increasing number of dimensions, with an asymptote of one. This indicates that in higher-dimensional spaces, the points seem to lie closer to the surface of a unit hypersphere around the origin. For both MSE and R^2, we do not observe any differences between the target spaces.

Let us now look at the results of the linear regression. It seems that for all the evaluation metrics, a two-dimensional target space yields the best result. With an increasing number of dimensions in the target space, performance tends to decrease. We can also observe that the amount of overfitting is optimal for a two-dimensional space and tends to increase with an increasing number of dimensions. A notable exception is the one-dimensional space which suffers strongly from overfitting and whose performance with respect to all three evaluation metrics is clearly worse than the baseline.

The optimal performance of a lasso regressor on the different target spaces yields similar results: For all target spaces, a certain amount of regularization can help to improve performance but too much regularization decreases performance. Again, we can only counteract a relatively small amount of the observed overfitting. As we can see in Table 3, again a two-dimensional space yields the best results. With respect to the optimal regularization factor β, we can observe that low-dimensional spaces with up to three dimensions seem to use larger values of β than higher-dimensional spaces with four dimensions and more. This difference in the degree of regularization is also reflected in the different degrees of overfitting observed for these groups of spaces.

Table 3 Performance of the zero baseline, the linear regression, and the lasso regression on target spaces of different dimensionality t derived with nonmetric SMACOF, along with the relative amount of overfitting. Best values for each column are highlighted for each of the regressors

Regressor	t	Test set performance			Amount of overfitting			β
		MSE	MED	R^2	MSE	MED	R^2	
Baseline	1	1.0000	**0.8664**	0.0000	1.0000	1.0000	1.0000	–
	2	1.0000	0.9580	0.0000	1.0000	1.0000	1.0000	–
	3	1.0000	0.9848	0.0000	1.0000	1.0000	1.0000	–
	4	1.0000	0.9956	0.0000	1.0000	1.0000	1.0000	–
	5	1.0000	0.9966	0.0000	1.0000	1.0000	1.0000	–
	6	1.0000	0.9973	0.0000	1.0000	1.0000	1.0000	–
	7	1.0000	0.9978	0.0000	1.0000	1.0000	1.0000	–
	8	1.0000	0.9980	0.0000	1.0000	1.0000	1.0000	–
	9	1.0000	0.9982	0.0000	1.0000	1.0000	1.0000	–
	10	1.0000	0.9984	0.0000	1.0000	1.0000	1.0000	–
Linear	1	1.1499	0.9046	−0.1499	59.0040	8.3419	−6.5413	–
	2	**0.4995**	**0.6370**	**0.5002**	38.9046	**6.5291**	1.9734	–
	3	0.5554	0.6979	0.4435	41.4309	6.7360	2.2243	–
	4	0.6086	0.7461	0.3706	42.4380	6.8305	2.6585	–
	5	0.6333	0.7692	0.3595	43.4577	6.9023	2.7405	–
	6	0.6359	0.7734	0.3469	43.4900	6.8770	2.8397	–
	7	0.6675	0.7956	0.3204	44.7364	6.9621	3.0741	–
	8	0.6846	0.8094	0.3033	45.1247	6.9876	3.2459	–
	9	0.6810	0.8078	0.2983	44.8367	6.9591	3.3004	–
	10	0.7107	0.8259	0.2807	46.0530	7.0432	3.5076	–
Lasso	1	0.9912	0.8368	0.0088	**1.3656**	**1.7043**	30.9878	1, 2
	2	**0.4728**	**0.6052**	**0.5271**	19.1298	4.5081	**1.8504**	0.02
	3	0.5322	0.6720	0.4722	19.4148	4.5725	2.0593	0.02
	4	0.5938	0.7316	0.3853	29.1237	5.6497	2.5413	0.005
	5	0.6180	0.7576	0.3755	35.1383	6.2167	2.6160	0.002
	6	0.6274	0.7651	0.3548	35.0732	6.1797	2.7724	0.001, 0.002
	7	0.6589	0.7839	0.3280	39.6352	5.0619	2.9979	0.001, 0.01
	8	0.6752	0.8022	0.3117	39.5496	6.5669	3.1527	0.001
	9	0.6680	0.7980	0.3108	38.8777	6.1359	3.1615	0.001, 0.002
	10	0.6993	0.8166	0.2924	35.3561	5.5563	3.3513	0.002, 0.005
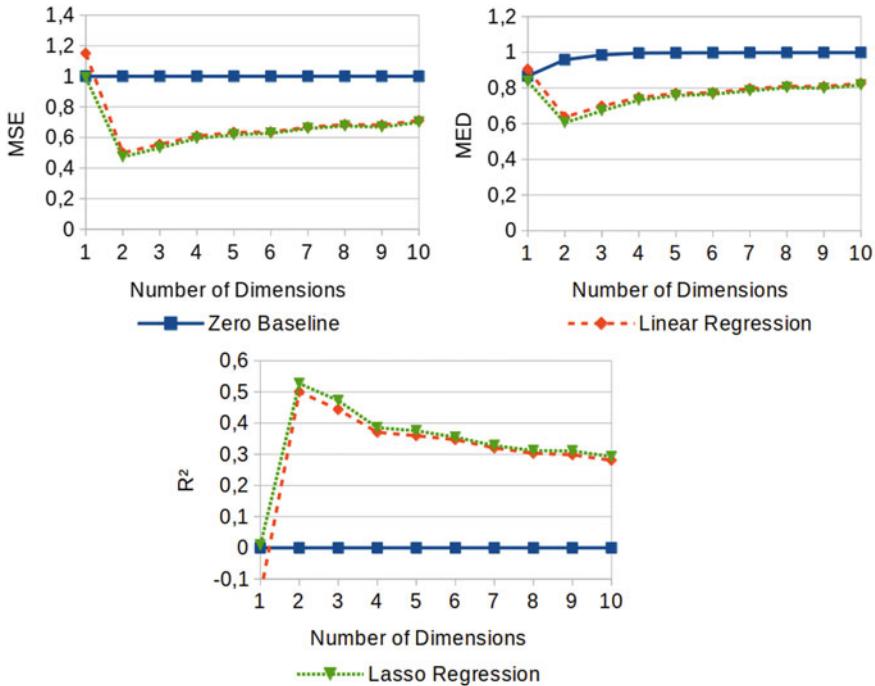
Fig. 6 Visualization of the regression results for MSE, MED, and R^2 as a function of the number of dimensions

Taken together, the results of our third experiment show that a higher-dimensional target space makes the regression problem more difficult, but that a one-dimensional target space does not contain enough semantic structure for a successful mapping. It seems that a two-dimensional space is in our case the optimal trade-off. However, even the performance of the lasso regressor on this space is far from satisfactory, urging for further research.

6 Conclusions

The contributions of this paper are twofold.

In our first study, we investigated whether the dissimilarity ratings obtained through SpAM are ratio scaled by applying both metric MDS (which assumes a ratio scale) and nonmetric MDS (which only assumes an ordinal scale). Both MDS variants produced comparable results—it thus seems that assuming a ratio scale is neither beneficial nor harmful. We therefore recommend to use nonmetric MDS as its underlying assumptions are weaker. Future studies on other data sets obtained through SpAM should seek to confirm or contradict our results.

In our second study, we analyzed whether learning a mapping from raw images to points in a psychological similarity space is possible. Our results showed that using the activations of a pre-trained ANN as features for a regression task seems to work in principle. However, we observed very strong overfitting tendencies in our experiments. Furthermore, the overall performance level we were able to achieve is still far from satisfactory. The results by Sanders and Nosofsky (2018) however show that larger amounts of training data can alleviate these problems. Future work in this area should focus on improvements in performance and robustness of this approach.

As follow-up work, we are currently conducting a study on a data set of shapes, where we plan to apply more sophisticated machine learning methods in order to counteract the observed overfitting tendencies.

References

Bechberger, L. (2020). lbechberger/LearningPsychologicalSpaces v1.3: Study on multidimensional scaling and neural networks on the NOUN dataset. https://doi.org/10.5281/zenodo.4061287.

Bechberger, L., & Kypridemou, E. (2018). Mapping images to psychological similarity spaces using neural networks. In *Proceedings of the 6th International Workshop on Artificial Intelligence and Cognition*.

Borg, I., & Groenen, J. F. (2005). *Modern multidimensional scaling: Theory and applications* (2nd ed.). Springer Series in Statistics. New York: Springer.

de Leeuw, J. (1977). Applications of convex analysis to multidimensional scaling. In *Recent development in statistics* (pp. 133–146). North Holland Publishing.

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

Gärdenfors, P. (2018). From sensations to concepts: A proposal for two learning processes. *Review of Philosophy and Psychology*, *10*(3), 441–464.

Goldstone, R. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers*, *26*(4), 381–386.

Horst, J. S., & Hout, M. C. (2016). The novel object and unusual name (NOUN) database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*(4), 1393–1409.

Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: A multidimensional scaling database with similarity ratings for 240 object categories from the massive memory picture database. *PLOS ONE*, *9*(11), 1–11.

Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient, spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General*, *142*(1), 256.

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1–27.

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*(2), 115–129.

Mitchell, T. M. (1997). *Machine learning*. McGraw Hill.

Pearson, K. (1895). Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, *58*(347–352), 240–242.

Peterson, J. C., Abbott, J. T., & Griffiths, T. L. (2017). Adapting deep network features to capture psychological representations: An abridged report. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17* (pp. 4934–4938).

Peterson, J. C., Abbott, J. T., & Griffiths, T. L. (2018). Evaluating (and improving) the corre-
    spondence between deep neural networks and human representations. *Cognitive Science*, *42*(8),
    2648–2669.
Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects
    in natural categories. *Cognitive Psychology*, *8*(3), 382–439.
Sanders, C. A., & Nosofsky, R. M. (2018). Using deep-learning representations of complex natural
    stimuli as input to psychological models of classification. In *Proceedings of the 2018 Conference
    of the Cognitive Science Society, Madison*.
Spearman, C. (1904). The proof and measurement of association between two things. *The American
    Journal of Psychology*, *15*(1), 72–101.
Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception
    architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision
    and Pattern Recognition* (pp. 2818–2826).
Wickelmaier, F. (2003). An introduction to MDS. *Sound Quality Research Unit, Aalborg University,
    Denmark*, *46*(5), 1–26.

Theories of Meaning for the Internet of Things

Michael Färber, Yulia Svetashova, and Andreas Harth

1 Why Traditional Knowledge Representation Is Insufficient

Future information systems, such as virtual assistants, augmented reality systems, and semi-autonomous or autonomous machines (Chan et al. 2009; Hermann et al. 2016), require access to large amounts of world knowledge in combination with sensor data. Consider a smart home scenario involving interconnected light bulbs. Here, a desired rule could be: "switch on the light in the hallway when somebody enters the home and set the light level in the hallway to below 50 lux." In this scenario, there needs to be a common understanding (i.e., *semantics*) of all the information (concepts and facts) mentioned in this command between the user and the device, such as "light," "hallway," "50 lux," but also of situational aspects, such as "when somebody enters the home." In a Health 2.0 scenario, connected devices measure parameters concerning a patient's health. The data need to be transformed (ideally automatically) into symbolically grounded knowledge and combined with the existing knowledge about health, diseases, and treatments (Henson et al. 2012).

M. Färber (✉)
Karlsruhe Institute of Technology (KIT), Institute AIFB, Karlsruhe, Germany
e-mail: michael.faerber@kit.edu

Y. Svetashova
Robert Bosch GmbH, Corporate Research and Advance Engineering, Robert-Bosch-Campus 1, 71272 Renningen, Germany
e-mail: svetashova@gmail.com

A. Harth
Friedrich-Alexander-University Erlangen-Nuremberg, Nuremberg, Germany
e-mail: andreas.harth@fau.de

Fraunhofer IIS-SCS, Nordostpark 84, 90411 Nuremberg, Germany

These examples demonstrate that knowledge representation for Internet of Things scenarios is needed. Specifically, on closer inspection, they indicate that three aspects are particularly essential for the Internet of Things knowledge representation:

1. How are perceptions and actions grounded and represented (in the Internet of Things terminology, sensors and actuators)?
2. How can machines and humans agree on a common understanding when referring to concepts and facts, and how can this common understanding be shared?
3. How can changes in the world be used in knowledge representation?

In the past, research on knowledge representation in computer science has mainly focused on developing and using static ontologies (i.e., as a *formal, explicit specification of a shared conceptualization in a domain of interest* (Studer et al. 1998)) and knowledge graphs (Fensel et al. 2020; Färber et al. 2018). Ontological languages, such as the *Resource Description Framework* (RDF) (Cyganiak et al. 2014), *RDF Schema* (Brickley and Guha 2014), and the *Web Ontology Language* (OWL) (Bechhofer et al. 2004), have been established to model parts of the world. To connect the world knowledge with sensor data, a few ad-hoc solutions have been proposed (e.g., Bonnet et al. 2000; Ganz et al. 2016, and Sect. 3). However, in our minds, all these technology is not capable of sufficiently incorporating the aspects of the Internet of Things as outlined above.

In this chapter, we want to take up the previous considerations on knowledge representation in the context of the Internet of Things; we thereby make use of content from epistemology—particularly, the semantic theories—for our discussion on an optimal knowledge representation, addressing research question 1 "How can we formally describe and model concepts?" outlined in Chap. 1 of this book. We can show that the problem of knowledge representation for the Internet of Things is by no means trivial and that questions about concrete implementations lead to fundamental questions of knowledge representation, such as the symbol ground problem (Harnad 1990) and the intersubjectivity problem (Reich 2010).

The topic of this chapter is highly interdisciplinary. Consequently, it is written for a diversity of user groups:

• Computer scientists, cognitive scientists, and IT practitioners who work on artificial intelligence systems and who are particularly interested in knowledge representation for the Internet of Things (e.g., designing ontologies for the Internet of Things);
• Philosophers who want to make autonomous systems and the Internet of Things accessible for their theories.

The chapter is structured as follows: After a detailed statement of the research problem in Sect. 1.1, we outline in Sects. 1.2 and 1.3 how our research problem is embedded in the scientific landscape of philosophy and computer science, respectively. In Sect. 2, we present a scenario in the Internet of Things context, which is used in the following sections to illustrate the concrete influences of theories of meaning on Internet of Things applications. Section 3 is dedicated to several semantic theo-

ries originating from philosophy and how they can be used to address our research problem. The chapter finishes with a summary in Sect. 4.

1.1 Problem Statement and Methodology

Problem Statement. The Internet of Things (IoT) refers to the idea of the "pervasive presence of a variety of things or objects around us—such as Radio-Frequency IDentification (RFID) tags, sensors, actuators, mobile phones, etc.—which, through unique addressing schemes, are able to interact with each other and cooperate with their neighbors to reach common goals" (Atzori et al. 2010). The Internet of Things has emerged as an important research topic and paradigm that can greatly affect a variety of aspects of everyday life. In the private setting, examples are smart homes, assisted living, and e-health. In the business setting, the Internet of Things is used, among other things, for automation and industrial manufacturing, logistics, and intelligent transportation.

We focus on the connection between the Internet of Things and knowledge representation. As such, we consider *intelligent agents*—defined as objects acting rationally (Russell and Norvig 2010) and often perceived as being identical to smart information systems—that

- have *sensors* (e.g., cameras, microphones, sensors for temperature and humidity) that allow them to perceive the environment,
- have *actuators* (e.g., displays, motors, light bulbs) that allow them to act in an environment,
- have an *interface* (e.g., buttons and dials, speech) to communicate with humans and
- can act semi-autonomously or autonomously.

In the future, humans and agents will increasingly co-exist side by side. For instance, humanoid robots with conversational artifical intelligence capabilities might become omnipresent. Moreover, agents will communicate with each other and thereby exchange knowledge to accomplish tasks in an autonomous way. However, obtaining a common understanding of the shared world and having the ability to refer to the same objects during communication is from an epistemological point of view nontrivial and by no means a matter of course. The crucial aspect in this context is the gap between the *represented world* (also called the *model*) and the *actual world* (see Fig. 1). It is related to mind-body dualism and specifically Descartes' mind-body problem in philosophy (Skirry 2006). Agents have access to the outside world (typically called *perception of the environment*) and are able to trigger changes in the world via actuators (i.e., they can change the outside world). This aspect is also related to the following questions: How can someone obtain the meaning of a text in a language unknown to him or her? How can someone interact with people without

the ability to speak the language of the people (see the Chinese room argument (Cole 2014))?

Methodology. We will outline the possibilities of modeling things for scenarios in the world of the Internet of Thing. Given the Internet of Things, an environment in which agents are situated with other agents, a *theory for knowledge representation on the Internet of Things* needs to

1. include sensors (perception) and actuators (action) in its formalization;
2. include the notion of multiple subjects (machines and humans) that interact with the environment and with each other (intersubjectivity);
3. include ways to describe changes in the represented world that should mirror changes in the real world, and vice versa (dynamics).

Acquiring the correct underlying foundations—and, in philosophical terms, the correct conditions of possibilities for acquiring and exchanging knowledge—is crucial to enabling the manifold benefits that arise from increased automation and human-computer interaction. As an example, let us take one of the prominent scenarios in the specific context of the Internet of Things—the so-called "onboarding" of devices. Onboarding is the process of connecting a sensor or a more complex Internet of Things device to the Internet and to a platform establishing an initial configuration and enabling services (Balestrini et al. 2017; Gupta and van Oorschot 2019). This process can either be automated or involves broad communities of device owners. In both cases, the problems of device-platform communication and deciding on identifiers (how to address a specific new device) require the acceptance of an adequate theory of meaning in the open context system. Such a system interacts with the changing world and needs to adapt accordingly.

This fact has already been noted by noteworthy philosophers and cognition scientists, such as (Gärdenfors 2000):

> When building robots that are capable of linguistic communication, the constructor must decide at an early stage how the robot grasps the meaning of words. A fundamental methodological decision is whether the meanings are determined by the state of the world or whether they are based on the robot's internal model of the world. (Gärdenfors 2000, p. 152)

Gärdenfors does not describe scenarios involving intelligent agents and does not show how the perception layer of a robotic system fits into his model of geometric spaces, which is the problem we address in this chapter. Specifically, we focus on *perception*, *multiple subjects*, and *world changes*.

1.2 Existing Solutions in Philosophy

In philosophy, the study of what knowledge is and how it can be represented (i.e., *epistemology*) and the study of how to acquire knowledge from an environment (i.e., *philosophy of perception*) are highly relevant to addressing the problem of knowledge
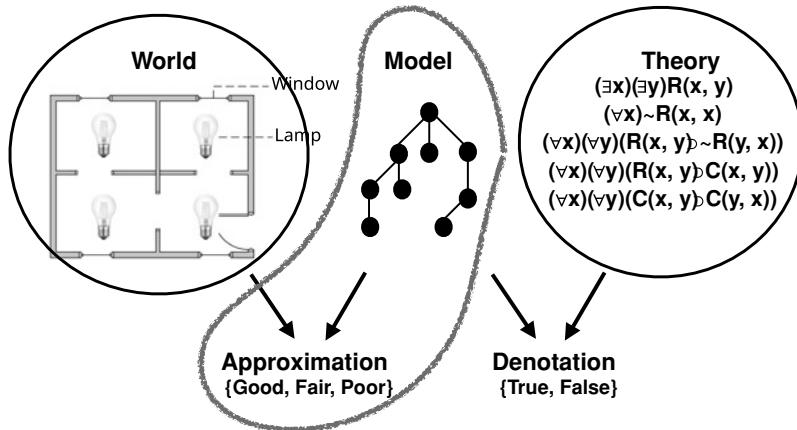
Fig. 1 Mediated reference theories distinguish between the world and a model of the world. Direct reference theories, in contrast, do not distinguish between the model and the world (i.e., the model *is* the world; illustration adapted from Sowa (2005))

representation for the Internet of Things. From these research areas, we can highlight the following aspects.

Theories of Meaning. Defining the meaning (particularly in the context of language also referred to as *semantics*) has always been an integral part of philosophy. In the 20th century, philosophy shifted its focus to language and the role of language in understanding. Particularly noteworthy is the groundbreaking work of Gottlob Frege (1848–1925), which can be seen as the basis for many achievements in the area of artificial intelligence. Frege's ideas come together in a mediated reference theory (see Fig. 1).

Frege challenged the belief that the meaning of a sentence directly depends on the meaning of its parts. The meaning of a sentence is its truth value and the meaning of its constituent expressions is their reference in the extra-linguistic reality. First, he explored the role of the proper names (which have direct reference) and concepts (which gain meaning only when their direct referent is specified). He then studied identity statements (in the form of $a = a$ or $a = b$) and came to the conclusion that direct reference theories do not adequately capture the meaning of identity statements. In particular, he pointed to the fact that the statements "Hesperus is the same planet as Hesperus" and "Hesperus is the same planet as Phosphorus" do not mean the same thing, even though the terms "Hesperus" and "Phosphorus" refer to the same extra-linguistic entity, the planet Venus. Thus, he came to an important distinction: the reference (*Bedeutung*) of a sentence is its truth value and the sense (*Sinn*) is the thought which it expresses. The questions that originated from Frege's arguments gave rise to many theories of meaning in logic and computer science and contoured the definition of meaning we accept in this chapter. Overall, Frege as a philosopher provided categories that other scientists questioned and developed.

We define meaning pragmatically as follows:

- *The meaning of symbols*. Following the idea that meaning is referential (see Gär-denfors 2000, pp. 151), the meaning of symbols (including names) are the objects to which they refer. For real-world entities, one can point directly towards the objects or mention their proper names. For abstract concepts (classes) and proper-ties, we take the Cartesian-Kantian two-world assumption as a basis and assume that, besides the real world we can see (*res extensa* or *phaenomenon*), there exists the world of ideas or thoughts (*res cogitans* or *noumenon*), and that classes and properties exist in this intellectual world.
- *The meaning of sentences*. Symbols can be arranged together to express statements (used synonymously to facts; sentences are the written counterpart). Statements bring us to a new level of meaning, the level of truthfulness. Each sentence can be *true* or *false*.

Theories of Truth. Given that statements can be true or false, questions of how statements stand in relation to the world and how statements can be tested concerning their truthfulness arise. Among the most commonly used theories of truth are as follow.

- *The correspondence theory of truth*: This theory can be considered the most basic theory of truth. True sentences capture the current state of affairs—objectively, without an observer. This theory is very much based on the actual world. The theory does not consider linking new knowledge to existing knowledge of a subject and differentiating between the varying knowledge levels of different people and how these people agree on the same meanings.
- *The coherence theory of truth*: This theory is coherence-centric and takes into consideration how new knowledge is incorporated into existing knowledge. State-ments are considered to be true if they are consistent with the statements (i.e., knowledge) obtained so far.
- *The consensus theory of truth (pragmatic theory of truth)*: This theory takes the different views of observers into consideration and is designed to align subjective views to other views. Statements are considered true if people (i.e., observers) agree on them.

It becomes immediately clear that these theories of truth do not exclude each other but rather have different foci. We argue that a comprehensive theory needs to take all of the theories' aspects into account. Particularly noteworthy is the fact that the theories focus mainly on knowledge and truth at a given point in time (see the construction of ontologies in computer science). Dynamic aspects, and thus the modeling of events, are insufficiently covered by these foundational theories.

1.3   Existing Solutions in Computer Science and Logic

In computer science and cognitive sciences, specifically the fields of knowledge representation and logic, the problem of how to represent knowledge about the world for Internet of Things scenarios has been addressed to some degree.

Theories of Meaning. In the past, computer scientists and logicians defined the meaning of objects in their knowledge representation models (e.g., ontologies) and methods for describing the world largely without an explicit connection to reality and perception. In particular, *model theory* (Tarski 1944) is the established way of defining the meaning of logic-based knowledge representation languages, such as the semantic web languages RDF, RDFS, and OWL.

Moreover, in the area of knowledge representation, it became popular to use *ontologies* (Staab and Studer 2010) and *knowledge graphs* (Fensel et al. 2020) as world models. Freely available open knowledge graphs form the Linked Open Data (LOD) cloud, which is used in various applications nowadays (Färber et al. 2018). However, since logic and model theory are very formal disciplines, there was no need to link knowledge representation to perception. Works on *ontology evaluation* and *ontology evolution* consider the process of creating and evaluating ontologies (in the sense used in computer science, i.e., as a formal model of a small domain of interest) as finding the lowest common denominator for modeling parts of the world. However, researchers mainly discuss common and best practices a team of developers can use to create an ontology. Early attempts at defining an ontology which incorporate temporal dynamics were made by Grenon and Smith (2004) and Heflin and Hendler (2000).

Overall, existing methods for modeling the world and defining meaning have the following drawbacks: (1) They disregard any explicit connection to reality. (2) They are omniscient and try to capture an (imposed) objective view of the world. (3) They are only able to express static knowledge but not changes in the world to a sufficient degree. In Sect. 1.2, we have carved out similar drawbacks regarding existing theories of meaning and truth.

If symbols are only identifiers, how can our minds create a link to an object in the real world (or in our conceptual worlds of ideas or thoughts)? How can we make sure that other subjects/minds have the same meaning; that is, link to the same object (e.g., when we only mention the object's identifier, such as http://dbpedia.org/resource/Karlsruhe or http://wikidata.org/entity/Q1040)? Is the meaning directly connected (grounded) to non-symbols? This problem is known as the *symbol grounding problem*: "How can you ever get off the symbol/symbol merry-go-round? How is symbol meaning to be grounded in something other than just more meaningless symbols?" (Harnad 1990). In the Semantic Web and Linked Data context, URIs are used as symbols for objects. The symbol-grounding problem is not often considered (Cregan 2007) or even solved. In particular, the aspects of perception, multiple subjects, and changes in the world—the focus in our chapter—for knowledge representation are not covered sufficiently. In the Internet of Things domain, we find only a

few works in this respect, such as the article by Hermann et al. (2017), who present grounded language learning in a 3D environment.

Theories of Truth. Theories of truth are traditionally proposed in philosophy. When we apply the theories of truth as introduced in Sect. 1.2 to the established and widely used semantic web technologies, such as RDF and OWL, and to knowledge representation ideas like knowledge graphs and linked open data, we can observe the following: (1) The RDF data model (Hayes and Patel-Schneider 2014) might fit to the *correspondence theory of truth* and to the *consensus theory* in the context of the Internet of Things. (2) Linked data can be regarded as an implementation of the *consensus theory* in the sense that data publishers and data consumers need to agree on common terms to use the linked data in a reasonable way. However, applications in the Internet of Things require more, since the (linked) data are subjected to changes over time and dependent on the perception (see, e.g., the sensor data from devices).

In recent years, approaches based on neural networks have been presented to represent entities and relations in knowledge graphs—as an implementation of a knowledge representation—in the form of vectors in a low-dimensional vector space (called *embeddings* Mikolov et al. 2013; Wang et al. 2017). Apart from the context of the entities and relations in the knowledge graph, external data sources have also been used to build these implicit knowledge representations. For instance, data from several modalities (text, images, speech, etc.) can be combined to form a unified, comprehensive representation in a low-dimensional space (Bruni et al. 2014). In the Internet of Things context, the representations are created based on sensor data, and thus, perceptions. We can argue that the formal method and technology of obtaining the sensory data (e.g., images, text, etc. of an object) and of transforming it into a common vector space (e.g., via machine learning techniques) has a direct influence on the meaning of objects or even constitutes the meaning itself.
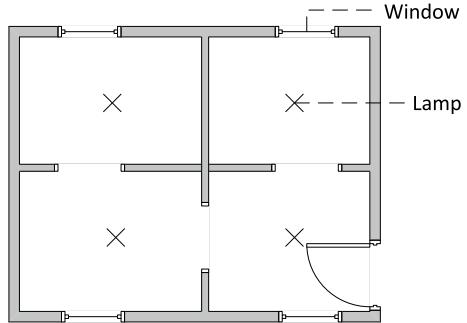
2   Motivating Scenario

In this section, we describe a smart home scenario, which will be used in the upcoming sections as an example of an Internet of Things scenario. It will show how the theories of meaning considered by us affect the way of modeling knowledge.

Consider the home of Alice (see Fig. 2) with four rooms: the hallway, the living room, the bathroom, and the bedroom. Each of the rooms is equipped with a light bulb that can be controlled via a network interface. Each room also has a window with controllable window blinds. Moreover, each room has a sensor to measure the light levels. The door has a sensor that detects when it is opened. A virtual assistant called Bob provides a user interface to the smart home via speech interaction. The more data and knowledge about the smart home is coupled with the virtual assistant, the more generic and flexible the virtual assistant needs to be.

Considering this scenario, we can point out several issues with respect to knowledge representation. The first issue concerns *naming*. Both Alice and Bob have to

Fig. 2 A smart home with various sensors (i.e., luminance sensors and door sensors) and actuators (i.e., network-controlled lamps and window blinds)

agree on the meaning of "the living room," so that Alice can ask Bob, "Is the light on in the living room?" Similarly, to affect a change in the world, Alice and Bob have to agree on names as references to objects, so that Alice is able to tell Bob to "switch off the light in the living room." A more elaborate command could be to "switch on the light in the hallway when somebody enters the home and set the light level in the hallway below 50 lux."

We assume that a shared understanding between the virtual assistant and the human user has to be configured when setting up the smart home (the so-called "onboarding problem"). The problem also arises when a new human user wants to interact with the smart home (e.g., when Carol visits Alice and wants to turn on the lights).

We can think of various other Internet of Things scenarios in which theories of meaning (also called *semantic theories*) become important for modeling the scenarios. For instance, in a Health 2.0 scenario as outlined by Henson et al. (2012), the sensor data gathered by Internet of Things devices need to be collected and transformed into symbolic information. This transformation allows the system to interpret the information and combine it with other existing, symbolically grounded knowledge (e.g., about diseases). Questions concerning the representation of perception, the inter-subjective agreement of concepts and facts, and the representation of dynamically changing knowledge arise.

3 Applying Theories of Meaning to the Internet of Things

Several theories of meaning have been proposed to link the real world with actual knowledge about it. In this section, we review the following semantic theories:

1. Model-theoretic semantics;
2. Possible world semantics;
3. Situation semantics;
4. Cognitive and distributional semantics.

These semantics have been chosen due to their popularity and "baselines" in previous work (Gärdenfors 2000, pp. 151). The first formalism is sometimes referred to as "extensional semantics" and the second formalism is referred to as "intensional semantics." Furthermore, some authors, such as Gärdenfors (2000), refer to "extensional semantics" instead of model theory and "intensional semantics" instead of "modal logic." Given the various and sometimes incompatible uses of "extensional" and "intensional" in the literature (Janas and Schwind 1979; Helbig and Glockner 2007; Lanotte and Merro 2018; Franconi et al. 2013), we use the terms "model-theoretic semantics" and "possible world semantics" for clarity.

In the following sections, we cover each theory of meaning in detail and apply it to the Internet of Things. Within each section, we first give a definition of the theory and outline its characteristics. Subsequently, we describe how the theory can be applied to model Internet of Things scenarios. We thereby focus primarily on the perception, intersubjectivity, and dynamics, because modeling these aspects is particularly crucial in the context of the Internet of Things (see Sect. 1).

3.1 Model-Theoretic Semantics for the Internet of Things

3.1.1 Definition and Current Use

Model-theoretic semantics can be encoded in various ways. In the following, we assume that the knowledge in embodied systems (e.g., a smart home) is described using sentences in first-order (predicate) logic. The meaning (i.e., the truth value) of the sentences is given via mapping to a world represented using set theory.

Extensional semantics is considered one of the realistic theories on semantics (Gärdenfors 2000). Expressions (names) are mapped to objects in the world (see the *theory of correspondence*). Predicates are then applied to a set of objects or relations between objects. Generally, using such a map, sentences can be assigned *true/false* values (see *truth conditions*). The "extension" of the sentence "Lassie is famous" is the logical value "true," since Lassie is famous. There is no anchoring of the language in a body (i.e., the meaning of words is modeled independently of individual subjects). This is known as the human capability of abstraction. All sentences being true constitute the world.

First-order predicate logic provides the foundation for formalizing current Semantic Web languages, such as RDF, RDFS, and OWL.

3.1.2 Application to the Internet of Things

While the languages with a formalization in model theory are mature and widely used, they do not cover the dimensions required in scenarios around the Internet of Things as outlined in the following:

Perception

The set-theoretic structure representing the world does not have any connection to the external world. Whether or not the term "Lassie" refers to Lassie the dog in the external world does not have any bearing on the truth value of the sentence. However, such a connection is needed to take perception (e.g., sensor data) in Internet of Things scenarios into account for modeling the world.

Intersubjectivity

The theory does not address the problem of reaching agreement on the meaning of terms across different agents. For instance, in the case of the semantic web languages RDF, RDFS, and OWL, there exists no defined mechanism that ensures different agents have the same notion of terms and sentences. Finding a shared understanding is left to the agents.

Dynamics

Traditional first-order predicate logic was developed to describe properties of things. That is, one can name things ("Lassie") and assign properties to them ("is famous"). The focus of such representations is to deduce new declarative sentences based on the given sentences. Some applications use first-order logic to represent events (e.g., "Lassie rescues the girl from drowning"), where the event ("rescuing") is treated as a property. While such representations might be suitable for some derivations, they do not cover the dynamics behind events sufficiently for scenarios in the Internet of Things.

Benefits and Limitations for the Internet of Things. The focus of model theory is to provide a notion of truth of sentences that allows for the specification of logical consequence. Logical consequence can help one check for satisfiability of sentences with regards to the world. It provides means to integrate data from multiple sources. However, model theory does not consider many aspects relevant in the Internet of Things, such as the connection of symbols and sentences to the real world or the question of how multiple agents can agree on the meaning of symbols. Furthermore, model theory lacks means to adequately formalize change, since the sentences are classically interpreted over a static model of the world.

3.2 Possible World Semantics for the Internet of Things

3.2.1 Definition and Current Use

The origins of possible world semantics can be traced back to Carnap (1947), Kripke (1959), and Montague (1974). Without loss of generality, we assume for the remaining part of the chapter that the possible world semantics are implemented via modal logic. More on the idea of possible worlds as the conceptual underpinning

of the modal logics can be found in Hughes et al. (1996) and Menzel (2017). In the following, we review the modal logic and its applicability for the Internet of Things scenarios.

With modal logic, expressions are mapped to a set of possible worlds, instead of a single world. Otherwise, the setting is the same as for the extensional semantics theory: sentences can have "truth conditions", and each proposition (sentence) has worlds in which it holds true.

To model these possible worlds, modal logic adds two new unary operators: $\square$ ("necessary") and $\lozenge$ ("possibly") to the set of Boolean connectors (negation, disjunction, conjunction and implication). The proposition is possible, if a world may exist in which this proposition is true. The proposition is necessary, if it has to be true in all worlds.

Dependent on the application context, modal operators can have different intuitive interpretations. For example, if one wants to represent temporal knowledge, $\square_{future} P$ may mean that proposition P is *always* true *in the future* and that $\lozenge_{future} P$ means P is *sometimes* true *in the future*. These different ways to interpret modal connectives give rise to various types of modal logics: tense, epistemic, deontic, dynamic, geometric, and others (see more in Goldblatt (2006)). Thus, they represent facts that are "necessarily/possibly" true, true "today/in the future", "believed/known" to be true, true "before/after an action", and true "locally/everywhere."

3.2.2 Application to the Internet of Things

We see many possibilities to use modal logic to capture the semantics in Internet of Things scenarios. As an example, Fig. 3 shows a system that interprets the voice input "Turn on the light" and acts differently depending on the location of the user. We can also consider such parameters as time of day and define different scenarios with temporal logics.

Modal logic as a kind of formal logic extends predicate logic by allowing it to express possibilities. Modal logic has mainly been used in formal sciences, such as logic (e.g., "ontology of possibilities"). However, it has not been applied extensively in computer science and, specifically, in Internet of Things contexts. We can observe that modal logic as an implementation of possible world semantics is better suited to the Internet of Things than model-theoretic semantics. However, modal logic is not perfectly suited for modeling knowledge of Internet of Things agents. This can be demonstrated by evaluating perception, intersubjectivity, and dynamics.

Perception

Similar to first-order predicate logic with a model-theoretic formalization, modal logic does not have any connection to the external world.

Intersubjectivity

Modal logic and its semantics are still based on a realistic idea (i.e., coordinating extra-linguistic entities to linguistics expressions). However, subjects' interpretations
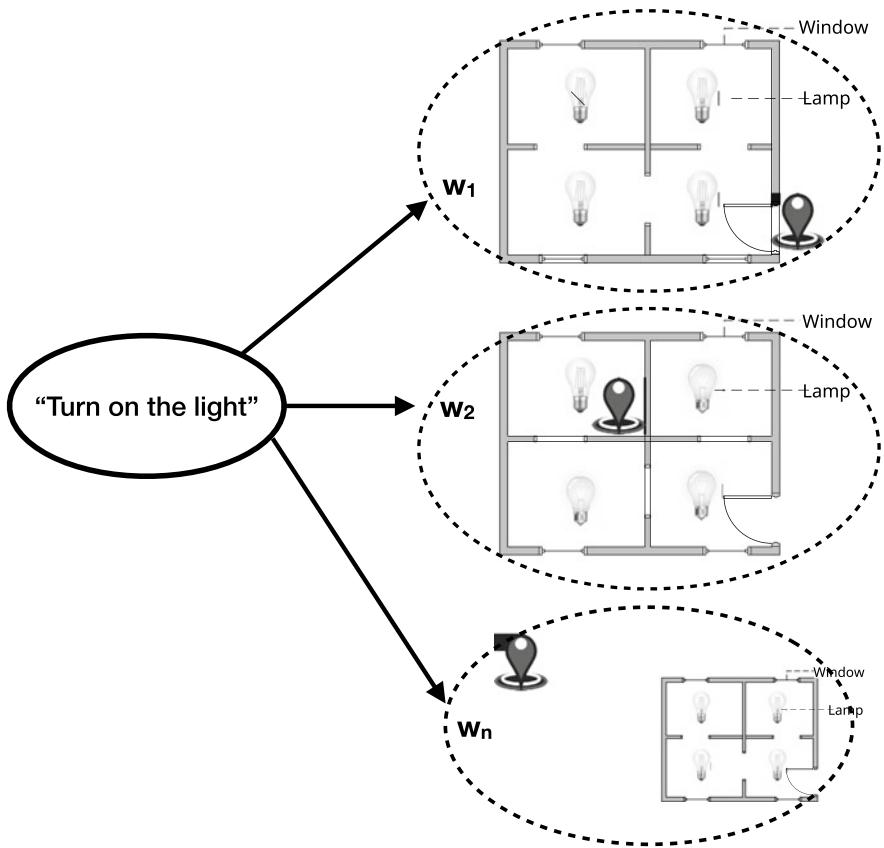
Fig. 3 Possible worlds in the smart home scenario

of the world can be represented as distinct worlds. In this way, modal logic allows us to model multiple worlds and to represent the knowledge of several agents (i.e., subjects).

Dynamics With the ability to add temporal operators, modal logic allows us to keep

track of states of resources over time. Furthermore, with the ability to keep track of state over time, one can detect events (i.e., state changes) and thus represent knowledge evolving over time.

Benefits and Limitations for the Internet of Things. The focus on logical consequence of sentences is one of the properties that possible world semantics shares with model-theoretic semantics. Neither has an explicit connection to the real world. Modal logic as an implementation of possible world semantics stands out from implementations of model-theoretic semantics by taking the aspects of intersubjectivity

and dynamics into account. Nevertheless, the possible world semantics only provide means to describe a changing world with sentences and to reason over such sentences, but not to actually affect changes in the world.

3.3   Situation Semantics for the Internet of Things

3.3.1   Definition and Current Use

The theory of situation semantics, another kind of realistic semantics, was developed by Jon Barwise and John Perry in their seminal book *Situations and Attitudes* (1983). In contrast to its predecessor *possible worlds semantics*, it postulates the principal of *partiality* of information available about the world. Limited parts of the world that are "clearly recognized […] in common sense and human language" and "can be comprehended as a whole in [their] own right" (Barwise and Perry 1980) are called *situations*. Situations stand in contrast to processes and activities. According to (Galton 2008):

> I believe that open processes and closed processes are very different kinds of things. The fact that we use the word 'process' for both of them perhaps lends some support to Sowa's use of this word as the most inclusive term, corresponding to what others have called situations or eventualities.

Devlin (2006), who formalized the basic notions of situation semantics and extended it to situation theory, emphasizes that information is always given "*about* some situation." It is constructed from discrete information units, called *infons*. An infon (σ) is a relational structure of shape, $\langle\langle R, a_1, \ldots, a_n, 0/1\rangle\rangle$, where R is an n-place relation, $a_1, \ldots, a_n$ are objects appropriate for the argument roles $i_1, \ldots, i_n$, and 0/1 are the *polarity* values indicating whether or not the objects $a_1, \ldots, a_n$ stand in the relation R.

Objects in the argument roles of an infon include individuals, properties, relations, space-time locations, situations, and parameters. *Parameters* in situation semantics act as variables (i.e., they reference arbitrary objects of a given type). To set parameters to concrete real-world entities, Barwise and Perry (1983) introduce an assignment mechanism called an *anchor*.

Unlike model-theoretic or possible worlds semantics, situation theory claims that an infon—roughly corresponding to a fact or statement—can be true (or false) only in the context of a particular situation. This relationship is written as $s \models \sigma$ (read as "*s* supports σ"), meaning that the fact represented by infon σ holds true in situation *s*.

Figure 4 shows an illustrative example of situation semantics for the Internet of Things scenario *smart home*. In this figure, we can see a limited part (*s*) of the world where we can distinguish several classes of objects: WindowBlind, Room, and LightBulb. Potentially, instances of these classes can be involved in many situations. One of them (i.e., TriggerBlindsUp) is that when it is dark in the room and already light outside in the morning, the window blinds are automatically raised
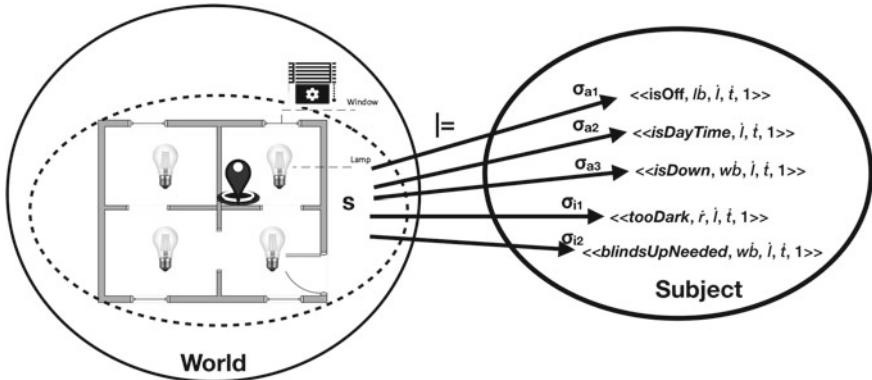
Fig. 4 `TriggerBlindsUp` situation in the smart home scenario

by the control system. We represent the relevant relations (*isDayTime*, *tooDark*, etc.) with the following infons where parameters $\dot{l}$ and $\dot{t}$ reference arbitrary spatial and temporal locations:

(σ_{a1}) $\langle\langle \text{isOff}, \dot{lb}, \dot{l}, \dot{t}, 1\rangle\rangle$, where parameter $\dot{lb}$ anchors objects of type `LightBulb`;

(σ_{a2}) $\langle\langle \text{isDayTime}, \dot{l}, \dot{t}, 1\rangle\rangle$;

(σ_{a3}) $\langle\langle \text{isDown}, \dot{wb}, \dot{l}, \dot{t}, 1\rangle\rangle$, with $\dot{wb}$ anchoring `WindowBlind` instances;

(σ_{i1}) $\langle\langle \text{tooDark}, \dot{r}, \dot{l}, \dot{t}, 1\rangle\rangle$, with $\dot{r}$ anchoring `Room` instances;

(σ_{i2}) $\langle\langle \text{blindsUpNeeded}, \dot{wb}, \dot{l}, \dot{t}, 1\rangle\rangle$, with $\dot{r}$ anchoring `WindowBlind` instances.

By using conjunction, disjunction, and anchoring, we can combine infons into more complex structures (i.e., *compound infons*). For situation `TriggerBlindsUp`, the infons form the compound infon: $s \models \sigma_{a1} \wedge \sigma_{a2} \wedge \sigma_{a3} \wedge \sigma_{i1} \wedge \sigma_{i2}$. The system that relies on this formalism can check whether these infons support the situation `TriggerBlindsUp` and use actuators to trigger the change in the real world.

Situation semantics distinguishes three types of situations: *utterance* situation (i.e., the immediate context of utterance, including a speaker and a hearer), *focal* situation (i.e., the part of the world referred to by the utterance), and *resource* situation (i.e., the situation used to support or to reason about focal or utterance situations (Devlin 2006)).

Meaning is acquired by linking utterances expressed in language to objects in the real world. This link, called the "speaker's connection" (Barwise and Perry 1983), determines the unique role of a subject in this theory: It is the agent who establishes such a link, and meaning is thus made relative to a specific agent. Figure 4 illustrates this possibly changing perspective. The subject perceives the room as dark: $\langle\langle \text{tooDark}, \dot{r}, \dot{l}, \dot{t}, 1\rangle\rangle$; one can imagine another subject for whom the polarity of the infon σ_{i1} would be 0.

In the area of the Internet of Things, certain information systems employ situation semantics as the core of their modeling of user behavior and sensor observations,

as well as the basis of context- and situation-awareness (see Heckmann et al. 2005; Kokar et al. 2009; Stocker et al. 2014, 2016). In the following, we will refer to these systems to show how situation semantics addresses the problems of perception, intersubjectivity, and representing dynamics.

3.3.2 Application to the Internet of Things

In the process of measurement, sensors transform signals of physical properties into numbers, thus generating numerical data. These data are challenging to store and manage and require near-instant access. The interpretation of the raw values requires modeling, finding patterns, and deriving abstractions. Abstractions reveal the properties of the observed real-world entities, show their dynamics, and place them into relations with their surroundings.

Perception

Sensor networks cannot perceive ("observe") situations directly; instead, as shown in Fig. 5, several components are needed to derive decisions and to take actions (see Kokar et al. 2009; Stocker et al. 2014). The process can be described as follows: The system takes sensor data as input, which then undergo the semantic enrichment process. Semantically annotated data is then transformed via a rule-based inference, digital signal processing, or machine learning algorithms into higher-level abstractions. These abstractions can be considered situations, which in turn can trigger actions and enable intelligent services. Both Stocker et al. (2014) and Kokar et al. (2009) exemplify how sensor input is transformed into a set of infons (called *observed* or *asserted* in Kokar et al. (2009)) and how new *inferred* infons are derived from them.

Situation semantics, therefore, works as a compliment to the algorithms that can directly process data generated in the perception layer. It is the way to organize sensory input in a task or goal-oriented environment. In addition, Stocker et al. (2014) argue that the persistence of situational knowledge in many cases is a desirable alternative to the persistence of sensor data and the key enabler of useful perceptual data in real time. Henson et al. (2012) describe an approach for deriving abstractions—essentially similar to situations—from sensory observations.
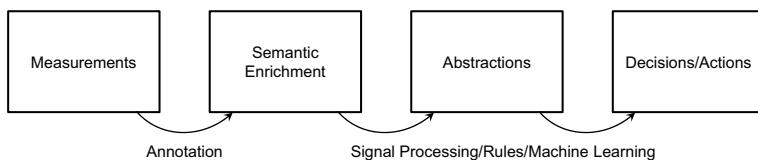

Fig. 5 Generic components of a system consuming sensor data

Intersubjectivity

In situation semantics, any relation between a real-world situation and its representation in a formal framework is relative to a specific subject. An agent recognizes or, in the terminology of Barwise and Perry (1983), "individuates" situations. Assigning values to certain parameters in the argument roles of an infon is always done by a particular subject. Situation semantics has an inherent mechanism to encode the subject's perspective, as well as to represent and to coordinate views of multiple subjects. The Internet of Things is often treated as a decentralized distributed system (Singh and Chopra 2017) where different agents generate situational knowledge individually. In this context, formalizing situation semantics can ease inter-agent communication and data integration (see the discussion in Stocker et al. (2014)).

Dynamics

Having *situation* as its central concept, situation semantics considers static representation of situations (as objects and their relations) and their dynamic aspect. According to Barwise and Perry (1983), "Events and episodes are situations in time … changes are sequences of situations." As a consequence, this theory has a built-in mechanism for representing temporal and spatial dynamics; namely, it introduces special types of objects that can fill argument roles of an infon (i.e., TIM, the type of a temporal location, and LOC, the type of a spatial location (Devlin 2006)). Thus, it is possible to represent whether a relation holds between the objects at a particular time in a particular location.

Stocker et al. (2014) use situation semantics to model observed situations in a road traffic scenario. By analyzing the road-pavement vibration data from three accelerometer sensors, they were able to detect vehicles in the proximity of sensing devices (*near*-relation) and their types (*light* or *heavy*). Observations, classified by the signal processing algorithms and modeled as sets of infons of shape, $\langle\langle \mathtt{near}, \mathtt{Vehicle}_x, l_x, t_x, 1\rangle\rangle$, enabled the inference of the velocity and the driving side of a vehicle via a custom set of rules. This example shows that this kind of representation is suitable for time-oriented data. Time-oriented data is a characteristic of most of the data generated in the Internet of Things (see more in Serpanos and Wolf (2017)).

Benefits and Limitations for the Internet of Things. Barwise and Perry were not the first to include situations as first-class citizens into a knowledge representation theory (see, e.g., situation calculus McCarthy 1963; McCarthy and Hayes 1969). Nevertheless, compared to its predecessors, situation semantics presents a richer formalism capable of representing higher level abstractions over raw sensor data, multiple viewpoints, and temporal-spatial dynamics.

Infons with their argument role structures can be reused across related situation types (e.g., how easy it will be to project the set of infons of the `TriggerBlindUp` situation to `TriggerBlindDown`). In many Internet of Things scenarios, the storage of raw data is not optimal due to the quantity and limitations of existing storage

solutions. Having a system as described in this section will allow us to store more meaningful and actionable pieces of information (situations) for certain signals to act upon in real time.

3.4 Cognitive and Distributional Semantics for the Internet of Things

3.4.1 Definition and Current Use

Cognitive semantics needs to be considered with respect to the general notion of *cognition*: instead of a subject perceiving the world with his senses and with language as the subject's ability to talk about the world, the focus is shifted to the mental representation of the world (i.e., to the subject's cognitive structures). Moreover, language becomes part of the cognitive structure. As such, concepts are elements in the subject's cognitive structure and without a direct reference to a reality. Thus, the meaning of concepts, etc., does not go beyond language, but is nothing else than using the language itself (see Ludwig Wittgenstein's theory of language) and therefore, the cognitive structures. These cognitive structures are subject to constant adaptation due to the interaction with the world. For instance, new concepts are learned and new findings are obtained. The world becomes *viable*. Overall, cognitive semantics is categorized as a non-realistic theory of semantics due to the exclusion of reality.

Focusing on the subject's cognitive structures, the question becomes what these cognitive structures look like and how they are created. Motivated by the biology of the human brain as the basis for any human's cognitive ability (Gärdenfors 2000, p. 257), neural networks and their mechanisms are typically considered the basis for cognition. Inputs, outputs, and internal representations of neural networks are modeled mathematically as geometrical (vector) spaces. Vector spaces are therefore used to represent things in the world, such as entities, concepts, and relations. Thus, knowledge is represented as *distributional representations* (e.g., embeddings) on a sub-symbolic level. *Meaning* is formalized as and reduced to a *distance function*. Similar objects tend to be spatially closer to each other in the vector space induced by the used neural network. Semantics is considered to be distributional (leading to the term *distributional semantics*), geometrical, and statistical.

Cognitive semantics and distributional semantics is not a new phenomenon: In 1954, Harris (1954) proposed that meaning is a function of distribution (see the famous quote: "a word is characterized by the company it keeps" (Harris 1954)). Contemporary philosophers and cognitive scientists use geometrical spaces to explain cognition and how concepts are formed by subjects. Gärdenfors (2000), for instance, considered the geometry of cognitive representations. In this cognitive space, points denote objects, while regions denote concepts (see Fig. 6 and book Chap. 2 for more information about Gärdenfors' cognitive framework).
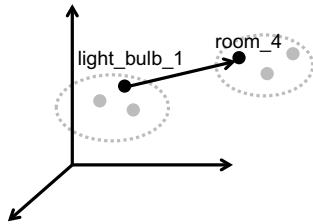
Fig. 6 Low-dimensional vector space representation in the smart home scenario with instances represented as points, concepts represented as areas, and predicates (relations) represented as vectors

Artificial neural networks have been used to simulate neural networks, and thereby, cognition. With the revival of research in artificial neural networks in recent years, research has been performed on how representations for terms, concepts, and predicates can be learned automatically (see, among other things, the approaches *TransE* and *TransH* (Wang et al. 2014)). The idea is to use the weights to the hidden layers of neural networks as representation (called *embeddings*). Guha (2015) proposed a model theory based on embeddings and adapted the Tarski model theory to embeddings.

In recent years, knowledge graph entities and relations (i.e., explicit knowledge representation formats) have also been embedded, showing that not only expressions can be represented in a distributed fashion, but also concepts and entities, as well as classes and relations. This allows us to model human cognition in a more natural way, because embeddings are learned for specific symbols.

3.4.2 Application to the Internet of Things

We assume that cognitive items, such as concepts, are represented in a sub-symbolic fashion, specifically, distributional semantics. Concepts are thus represented in a geometrical space. We use neural-network-based embedding methods as concrete implementation for distributional semantics. Figure 6 shows an example of representing items for the smart home scenario. Distributional semantics is amenable to modeling perception, intersubjectivity, and dynamics in the following respect:

Perception

Distributional semantics differs (with respect to perception) from other semantic theories in several ways:

- The meanings of concepts and facts are represented in a distributed fashion, not as singular units or symbols. In the smart home scenario, specific light bulbs and rooms are encoded as embedding vectors (see Fig. 6).
- The representations and, thus, the meanings of concepts are not static, but can be subject to constant change. To overcome this issue, time-dependent embeddings can be learned (Nguyen et al. 2018). In the smart home scenario, agents can learn

the embeddings of the different light bulbs and the embedding space of light bulbs per se based on the sensor data used as input for a neural network.

- A symbol grounding is possible as long as some form of input data (i.e., sensor data) is provided (and does not change abruptly).
- There are indications that human cognition aggregates the perceptions of different modalities of one unit (e.g., concept or concrete entity). For instance, the image of a dog and the sound of a dog are immediately perceived as belonging to the same unit. The same phenomenon can be observed when a multilingual person switches between languages whilst referring to the same concepts. As is the case with embedding methods from machine learning, such a fusion of sensor data from different modalities is possible. In the Internet of Things context, an embedding vector can be learned jointly based on different modalities.

Overall, perception is reduced to learning embeddings.

Intersubjectivity

Talking about and reaching an agreement on expressions between several agents can be traced back to using the same learned representations (i.e., embedding vectors) and the same conceptual structures (i.e., the distributional space). Even if different initialization values for the embedding spaces are given, subjects can use the same learning function to learn the same concepts. In the smart home scenario, the agents might differ in the exact points of the single light bulbs and rooms, since they rely on their own embedding learning and usage. However, they can agree on the same instances and concepts if the embeddings share the same characteristics (e.g., having nearly the same distances to other embeddings in the vector space). Overall, learning representations and meaning are reduced to learning and applying the same mathematical functions and models.

Dynamics

Describing changes in the world, such as events, is not sufficiently possible in the cognitive theory of semantics. If embeddings as distributed representations are learned or adapted online (i.e., in a permanent fashion, not only once at the beginning), then changes in the world may change the embeddings. However, the change itself is not represented. In the smart home scenario, an event might be light bulb number 4 switching on. The concepts involved in this event, such as light bulb #4, the positions of the light bulbs, and room #1 remain the same.

Benefits and Limitations for the Internet of Things. A characteristic of cognitive/distributional semantics is that information, such as concepts and facts, is not represented in the form of symbols, but in a sub-symbolic fashion as points and spaces in a vector space. This allows a more continuous distance function and an agreement on concepts and facts in the world as a continuous process. Talking about and reaching an agreement on expressions between several subjects can be traced back to using the same learned representations (i.e., embedding vectors) and the same conceptual structures (i.e., the distributional space). Thus, distributional semantics is heavily based on mathematics, which benefits the modeling of data in the Internet

of Things setting. However, describing changes in the world, such as events, is not sufficiently possible in the cognitive theory of semantics.

4   Conclusion

In this chapter, we have considered the theoretical foundations for representing knowledge in the Internet of Things context. Based on the peculiarities of the Internet of Things, we have outlined three dimensions that must be examined with respect to theories of meaning:

1. *Perception*: How can a theory of meaning incorporate "direct access" to the world (e.g., via sensors)?
2. *Intersubjectivity*: How can the world view of several subjects (i.e., agents in the Internet of Things) be modeled coherently?
3. *Dynamics*: How can the change of knowledge be modeled sufficiently? Which aspects of time can be represented?

We considered the following theories of meaning:

1. The model theory (extensional semantics)
2. Modal logic (intensional semantics)
3. Situation semantics
4. Cognitive/distributional semantics.

The single theories have the following advantages and disadvantages (see also Table 1):

1. *Model-theoretic semantics* is the simplest model in our series of considered semantic theories. This semantic theory can be used to formulate sentences and their truth values. However, it does not provide us with techniques or formalisms for modeling reality to the highest degree (i.e., with its unstable and experiential nature).

Table 1   Overview of how the challenges of perception, intersubjectivity, and dynamics are met by the various theories of semantics

	Model-theoret. semantics	Possible world semantics	Situation semantics	Conceptual/distrib. semantics
Perception			✓ (indirectly, via sensor data processed by machine learning)	✓
Intersubjectivity		✓	✓	✓ (communicating the dimensions)
Dynamics		✓	✓	

2. *Possible world semantics* also does not provide us with an explicit connection to the real world that we could use to adequately model knowledge in the Internet of Things domain. However, it allows implementations for modeling temporal modalities and agent's beliefs. As such, this semantic theory first lays theoretical foundations concerning dynamics and intersubjectivity. However, the foundational questions of how agents reach an agreement and how this fact about the agreement can be represented are not covered.

3. *Situation semantics* can be considered another layer in the pyramid of knowledge representation formalisms, as raw sensor data, multiple viewpoints, and temporal-spatial dynamics can be represented to some degree. However, we believe that this formalism is also not the optimal semantic theory for Internet of Things scenarios, as it leaves too many questions unanswered, particularly concerning perception and intersubjectivity.

4. *Cognitive and distributional semantics* can be judged in a manner similar to situation semantics when it comes to Internet of Things applications. Compared to the previous semantic theories, cognitive and distributional semantics are rather empirical (i.e., data-driven theories). The introduction of different levels of cognition and the fact that symbolic knowledge representation can be connected to sub-symbolic knowledge representation is appealing, particularly when it comes to data gathered by sensors. Intersubjectivity can be reduced to empirical training using data and mathematical functions (encoded in the form of [neural] networks). We see the main lack of this semantic theory in the (elegant) modeling of knowledge change over time.

Overall, we came to the conclusion that each of the semantic theories helps in modeling specific aspects, while not sufficiently covering all three aspects simultaneously. For the future, working on the advancements of situational semantics and distributional semantics and combining them towards a united semantic theory can be very fruitful for developing future intelligent information systems.

References

Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer Networks*, *54*(15), 2787–2805.

Balestrini, M., Seiz, G., Peña, L. L., & Camprodon, G. (2017). Onboarding communities to the IoT. In *International Conference on Internet Science*, INSCI (Vol. 17, pp. 19–27). Springer.

Barwise, J., & Perry, J. (1980). The situation underground. In J. Barwise, I. Sag, & Stanford Cognitive Science Group (Eds.), *Stanford Working Papers in Semantics* (Vol. 1, pp. 1–55). Stanford University Press.

Barwise, J., & Perry, J. (1983). *Situations and attitudes*. Cambridge (Mass.): MIT Press.

Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., & Stein, L. A. (2004). OWL web ontology language reference. Retrieved May 03, 2020, from https://www.w3.org/TR/owl-ref/.

Bonnet, P., Gehrke, J., & Seshadri, P. (2000). Querying the physical world. *IEEE Personal Communications*, *7*(5), 10–15.

Brickley, D., & Guha, R. (2014). RDF Schema 1.1. Recommendation, W3C. http://www.w3.org/TR/2014/REC-rdf-schema-20140225/. Latest version available at http://www.w3.org/TR/rdf-schema/.

Bruni, E., Tran, N., & Baroni, M. (2014). Multimodal distributional semantics. *Journal of Artificial Intelligence Research*, *49*, 1–47.

Carnap, R. (1947). *Meaning and necessity: A study in semantics and modal logic*. University of Chicago Press.

Chan, M., Campo, E., Estève, D., & Fourniols, J.-Y. (2009). Smart homes-current features and future perspectives. *Maturitas*, *64*(2), 90–97.

Cole, D. (2014). The Chinese room argument. Stanford Encyclopedia of Philosophy. Retrieved September 02, 2019, from https://plato.stanford.edu/entries/chinese-room/.

Cregan, A. (2007). Symbol grounding for the semantic web. In *Proceedings of the 4th European Semantic Web Conference*, ESWC'07 (pp. 429–442).

Cyganiak, R., Wood, D., & Lanthaler, M. (Eds.). (2014). RDF 1.1 concepts and abstract syntax. Retrieved May 10, 2020, from http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/.

Devlin, K. (2006). Situation theory and situation semantics. In *Handbook of the history of logic* (Vol. 7, pp. 601–664). Elsevier.

Färber, M., Bartscherer, F., Menne, C., & Rettinger, A. (2018). Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. *Semantic Web*, *9*(1), 77–129.

Fensel, D., Simsek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., et al. (2020). *Knowledge graphs: Methodology, tools and selected use cases*. Springer.

Franconi, E., Gutiérrez, C., Mosca, A., Pirrò, G., & Rosati, R. (2013). The logic of extensional RDFS. In *Proceedings of the 12th International Semantic Web Conference*, ISWC'13 (pp. 101–116).

Galton, A. (2008). Experience and history: Processes and their relation to events. *Journal of Logic and Computation*, *18*, 323–340.

Ganz, F., Barnaghi, P. M., & Carrez, F. (2016). Automated semantic knowledge acquisition from sensor data. *IEEE Systems Journal*, *10*(3), 1214–1225.

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

Goldblatt, R. (2006). Mathematical modal logic: A view of its evolution. In *Handbook of the history of logic* (Vol. 7, pp. 1–98). Elsevier.

Grenon, P., & Smith, B. (2004). SNAP and SPAN: Towards dynamic spatial ontology. *Spatial Cognition & Computation*, *4*(1), 69–104.

Guha, R. V. (2015). Towards a model theory for distributed representations. In *Proceedings of the 2015 AAAI Spring Symposia* (pp. 22–26).

Gupta, H., & van Oorschot, P. C. (2019). Onboarding and software update architecture for IoT devices. In *Proceedings of the 17th International Conference on Privacy, Security and Trust*, PST'19 (pp. 1–11).

Harnad, S. (1990). The symbol grounding problem. *Physica D*, *42*, 335–346.

Harris, Z. S. (1954). Distributional Structure. *Word*, *10*(2–3), 146–162.

Hayes, P., & Patel-Schneider, P. (2014). RDF 1.1 semantics. Recommendation, W3C. http://www.w3.org/TR/2014/REC-rdf11-mt-20140225/. Latest version available at http://www.w3.org/TR/rdf11-mt/.

Heckmann, D., Schwartz, T., Brandherm, B., Schmitz, M., & von Wilamowitz-Moellendorff, M. (2005). GUMO: The general user model ontology. In *Proceedings of the 10th International Conference on User Modeling*, UM'05 (pp. 428–432).

Heflin, J., & Hendler, J. A. (2000). Dynamic ontologies on the web. In *Proceedings of the 17th National Conference on Artificial Intelligence and 12th Conference on Innovative Applications of Artificial Intelligence* (pp. 443–449).

Helbig, H., & Glockner, I. (2007). The role of intensional and extensional interpretation in semantic representations—The intensional and preextensional layers in MultiNet. In *Proceedings of the 4th International Workshop on Natural Language Processing and Cognitive Science*, NLPCS'07 (pp. 92–105).

Henson, C. A., Sheth, A. P., & Thirunarayan, K. (2012). Semantic perception: Converting sensory observations to abstractions. *IEEE Internet Computing*, *16*(2), 26–34.

Hermann, K. M., Hill, F., Green, S., Wang, F., Faulkner, R., Soyer, H., Szepesvari, D., Czarnecki, W. M., Jaderberg, M., Teplyashin, D., Wainwright, M., Apps, C., Hassabis, D., & Blunsom, P. (2017). Grounded Language Learning in a Simulated 3D World. CoRR. arXiv:1706.06551.

Hermann, M., Pentek, T., & Otto, B. (2016). Design principles for Industrie 4.0 scenarios. In *Proceedings of the 49th Hawaii International Conference on System Sciences*, HICSS'16 (pp. 3928–3937).

Hughes, G. E., Cresswell, M. J., & Cresswell, M. M. (1996). *A new introduction to modal logic*. Psychology Press.

Janas, J. M., & Schwind, C. B. (1979). Extensional semanntic networks: Their representation, application, and generation. In N. V. Findler (Ed.), *Associative networks* (pp. 267–302). Academic Press.

Kokar, M. M., Matheus, C. J., & Baclawski, K. (2009). Ontology-based situation awareness. *Information Fusion*, *10*(1), 83–98.

Kripke, S. A. (1959). A completeness theorem in modal logic. *The Journal of Symbolic Logic*, *24*(1), 1–14.

Lanotte, R., & Merro, M. (2018). A semantic theory of the internet of things. *Information and Computation*, *259*(1), 72–101.

McCarthy, J. (1963). *Situations, actions, and causal laws*. Technical report, Stanford University, Department of Computer Science.

McCarthy, J., & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer, & D. Michie (Eds.), *Machine intelligence* (Vol. 4, pp. 463–502). Edinburgh University Press.

Menzel, C. (2017). Possible worlds. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University, Winter 2017 edition.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of the 27th Annual Conference on Neural Information Processing Systems*, NIPS'13 (pp. 3111–3119).

Montague, R. (1974). *Formal philosophy: Selected papers of Richard Montague*. Yale University Press.

Nguyen, G. H., Lee, J. B., Rossi, R. A., Ahmed, N. K., Koh, E., & Kim, S. (2018). Continuous-time dynamic network embeddings. In *Proceedings of the 2018 World Wide Web Conference*, WWW'18 (pp. 969–976).

Reich, W. (2010). Three problems of intersubjectivity-and one solution. *Sociological Theory*, *28*(1), 40–63.

Russell, S. J., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (3rd ed.). Pearson Education.

Serpanos, D., & Wolf, M. (2017). *Internet-of-things (IoT) systems: Architectures, algorithms, methodologies*. Springer.

Singh, M. P. & Chopra, A. K. (2017). The internet of things and multiagent systems: Decentralized intelligence in distributed computing. In *Proceedings of the 37th IEEE International Conference on Distributed Computing Systems*, ICDCS'17 (pp. 1738–1747).

Skirry, J. (2006). The mind-body problem. Retrieved May 05, 2020, from https://www.iep.utm.edu/descmind/#H4.

Sowa, J. (2005). Theories, models, reasoning, language, and truth. Retrieved August 31, 2019, from http://www.jfsowa.com/logic/theories.htm.

Staab, S., & Studer, R. (2010). *Handbook on ontologies*. Springer Science & Business Media.

Stocker, M., Nikander, J., Huitu, H., Jalli, M., & Koistinen, M. (2016). Representing situational knowledge for disease outbreaks in agriculture. *Agrárinformatika/Journal of Agricultural Informatics*, *7*(2), 29–39.

Stocker, M., Rönkkö, M., & Kolehmainen, M. (2014). Situational knowledge representation for traffic observed by a pavement vibration sensor network. *IEEE Transactions on Intelligent Transportation Systems*, *15*(4), 1441–1450.

Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, *25*(1–2), 161–197.

Tarski, A. (1944). The semantic conception of truth: And the foundations of semantics. *Philosophy and Phenomenological Research*, *4*(3), 341–376.

Wang, Q., Mao, Z., Wang, B., & Guo, L. (2017). Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, *29*(12), 2724–2743.

Wang, Z., Zhang, J., Feng, J., & Chen, Z. (2014). Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, AAAI'14 (pp. 1112–1119).

A Qualitative Similarity Framework for the Interpretation of Natural Language Similarity Expressions

Helmar Gust and Carla Umbach

1 Introduction

In this paper, a representational framework is presented featuring a qualitative notion of similarity. It is aimed at issues of natural language semantics, in particular the semantics of expressions of similarity and sameness and their role in comparison and ad-hoc kind formation.[1] Starting point was the interpretation of such expressions in German and English, for example *so/such*, *ähnlich/similar*, and *gleich/same*, which all denote similarity in some sense. It would be unsatisfactory, however, to treat similarity as a primitive predicate because semantic differences between individual similarity expressions would be obscured, for example, the fact that *ähnlich/similar* are gradable while *so/such* and *gleich/same* are not (see Umbach and Gust in print). Furthermore it would be difficult to establish the connection between similarity expressed by scalar and non-scalar equative comparison constructions, as shown in (1).

---

[1] The notion of *kinds* in linguistics is closely connected to the notion of *concepts* in psychology (Carlson 2010). Moreover, *ad-hoc categories* formed by linguistic expressions show core characteristics of concepts (Barsalou 1983). We thus assume that kinds formed ad-hoc by similarity expressions closely correspond to concepts, see Umbach and Stolterfoht (in prep).

H. Gust
Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany
e-mail: hgust@uos.de

C. Umbach (✉)
Department of German Language & Literature I, University of Cologne, Cologne, Germany
e-mail: carla.umbach@uni-koeln.de

(1) a.   Anna is as tall as Berta.   scalar/adjectival
     b.   Anna has a car like Berta's.   non-scalar/nominal
     c.   Anna is dancing just like Berta.   non-scalar/verbal

Finally, a primitive similarity predicate would leave no room to account for the observation that certain similarity expressions, in certain contexts, can be used to form ad-hoc kinds. German *so* as well as English *such* combined with nominal expressions may refer to kinds (or concepts) instead of individuals. In (2a, b), for example, *so ein Fahrzeug/such a vehicle* does not refer to a particular vehicle but instead to an ad-hoc created kind of vehicles including the set of vehicles similar to the one the speaker points to. Umbach and Stolterfoht present experimental exidence that features licensing ad-hoc kinds must be principally connected to concepts, excluding factual and statistical properties (König and Umbach 2018; Umbach and Gust 2014; Umbach and Stolterfoht in prep.). Thus, a complex notion of similarity not only provides a detailed semantic interpretation of natural language similarity expressions—it opens a window into mechanisms of concept formation.

(2) (Speaker points to an oversized car that makes enormous noise:)
     a.   So ein Fahrzeug wird in den Innenstädten bald verboten sein.
     b.   Such a vehicle will soon be banned in the inner cities.

The framework in this paper offers a way to spell out the notion of similarity in some detail without being forced to leave the well-established ground of referential semantics. The core idea is to make use of attribute spaces representing complex features of individuals, and to make use of predicates defined on such features determining the granularity of representation. In accordance with referential semantics we assume that natural language expressions refer to entities, or categories of entities, in the real world. However, access is only indirect, mediated by *generalized measure functions* mapping real world entities to points in attribute spaces (this is called a *mediated reference theory* in Färber, Svetashova and Harth, this volume). Similarity is a key concept in our framework because it provides a variable notion of *identity/indistinguishability with respect to a representation*: Individuals count as similar if their features in a particular attribute space, given a particular granularity, cannot be distinguished.

This system provides a powerful and flexible tool in the analysis of natural language semantics facilitating detailed interpretations of similarity expressions (*so*, *such*, *similar* etc.). Beyond, and maybe even more relevant, this system offers the possibility to analyze linguistic ad-hoc kind formation constructions, for example, by *so/such* demonstratives and equative comparison as in (1) and (2). It is important to realize, however, that this system is basically a multidimensional generalization of degree semantics (e.g., Kennedy 1999) complemented by a method for varying granularity. From this point of view, our framework is anchored in referential semantics just as much as degree semantics is.

Attribute spaces are well-established methods of representation in AI[2] and also in some branches of natural language semantics, e.g., in frame-based approaches (Barsalou 1992; Minsky 1975). What distinguishes attribute spaces and representations as proposed in this paper from classical frame-based approaches is that we focus on systems of predicates on points in attribute spaces in contrast to the points in these spaces themselves, thereby introducing a qualitative aspect, for instance in modelling comparison. This idea is connected to the idea of micro-theories (see, e.g., in Cyc[3] or other ontology languages) which talk about small parts of the world covered, e.g., by a single concept like *chair*, *vehicle*, *elephant*, *human,* etc., but also about actions and events. We expect that such micro-theories provide some kind of prototypes or exemplars, positive and also negative ones. Maybe we just imaginate such exemplars. Here is a typical way how to introduce the concept of a physical object in a beginners lecture in experimental physics by imagination of a positive example[4]: "Think of a red steel ball of ten centimeters diameter in front of you. It need not to be red, it need not to be made from steel, it need not have a diameter of ten centimeters and it need not be a ball." This shows that even abstract concepts can be characterized by exemplars (real or imagined) together with the specification of relevant dimensions in an attribute space.

This paper is structured in the following way: In Sect. 2 we develop a formal theory of representation making use of predicate systems over attribute spaces. Section 3 gives a brief overview over the interpretation of natural language similarity expressions and the role of similarity in ad-hoc kind formation and equative comparison. Since the focus of this paper is on formal characteristics of the representational framework, we will not go into linguistic details.[5] In Sect. 4 we develop a formal similarity concept based on methods provided in Sect. 2. Section 5 shows how to use granularity and hierarchies of representations in order to model gradabilty along non-scalar dimensions.

2 Representations in Multi-dimensional Attribute Spaces

We start from the idea that natural language expressions refer to entities or categories (or even higher order structures, e.g., relations) of entities in the real world, but in an indirect way. Access to these entities or categories is mediated by a function we call *generalized measure function*, e.g., $car_1 \Rightarrow$ {horse_power: 100 ps, weight: 1680 kg,

---

[2]Starting from Minsky's frames (Minsky 1975) and feature structures, up to modern approaches based on description logics (for an overview see https://en.wikipedia.org/wiki/Description_logic).

[3]For micro-theories in Cyc see, e.g., https://pdfs.semanticscholar.org/4f28/6fdf9280449588b9d3781c9c897da28e0cff.pdf.

[4]For an overview of the imagery debate see https://plato.stanford.edu/entries/mental-imagery/.

[5]Readers primarily interested in formal frameworks might skip Sect. 3. Readers primarily interested in semantics might want to start with Sect. 3 and eventually go back.

color: green …}. This is related to what is called *observables* in physics[6]: Such a function assigns observable attributes (elements of an attribute space) to entities or classes of entities in the world.[7] The referential power of language predicates like *car* (their meaning in the world) can thus be approximated by classifiers. Such classifiers should be effectively computable characteristic functions of predicates.[8] They operate on attribute spaces (or higher order structures based on attribute spaces).[9] Still, we can go back from predicates on points in attribute spaces to predicates on the entities in the world via the inverse image of the generalized measure functions.

On the worldy side, a domain includes a set of relevant predicates P talking about entities in the world. According to the notion of a representation in this paper, these predicates have counterparts on the representational side marked by a star (*) in Fig. 1. Counterpart predicates are required to be consistent with their originals; more precisely, they have to agree in truth value on the set of positive and negative exemplars of the original predicate. Moreover, counterpart predicates will be assumed to have convex extensions. As a consequence, they must be true on all points in the convex closure of the images of the positive exemplars (see Fig. 1 below). In addition, we stipulate that the extensions of counterpart predicates must be open[10] in some given topology on attribute spaces. This ensures that small changes in the representation (in the sense of the given topology) do not change the truth-values of these predicates.

---

[6]There is a long-standing debate about the dichotomy of observables vs. theoretical terms in philosophy, see https://plato.stanford.edu/entries/theoretical-terms-science/. We take a naive view here: observables are functions assigning values to entities in the world which can be determined by 'simple' measurements. Examples are *temperature*, *length*, *width*, *height*, *color*, *position*, etc., in contrast to values for energy (which in case of heat, for example, depends on temperature, mass and specific heat of the matter).

[7]Our approach is non-constructive since we do not construct representations, but instead have systems of constraints which representations must obey. Bechberger and Kühnberger (this volume) discuss approaches for learning feature space representations by multidimensional scaling. They optimize these representations by using artificial neural networks. From our point of view, they try to learn a feature space F and a measure function μ from similarity and dissimilarity judgments of subjects. In this case, μ maps stimuli (elements of a stimuli domain D) to points in F.

Their approach is restricted such that all dimensions of F have a uniform structure. Essentially F is an euclidean vector space in their approach. There is no canonical interpretation of the dimensions found, and therefore, no link to natural language expressions. In a second step, the goal is to find classifiers which approximate meaningful subclasses of the stimuli space, which may then lead to interpretations of the dimensions. Bechberger and Kühnberger discuss this as a quality measure suited in determining the number of dimensions of F. They generalize the approach to handle unseen stimuli.

[8]Classification problems are common in artificial intelligence, where classifiers are trained on huge example sets to be able to classify unseen examples without error. Analogous to our approach, the first step is to find a suitable representation of the real world problems which can be handled by the classification algorithm. Then the example cases have to be translated into this representation in order for the classifier to be able to learn.

[9]We may want to restrict computational complexity of classifiers since there should be efficient algorithms for classification. We will pay with accuracy to get easy to classify areas within the attribute space.

[10]Open sets are sets without a border. Think of a ball in three-dimensional Euclidean space as something like a tomato: It has a crisp border. If we remove the border by peeling, it is unclear where the tomato ends.
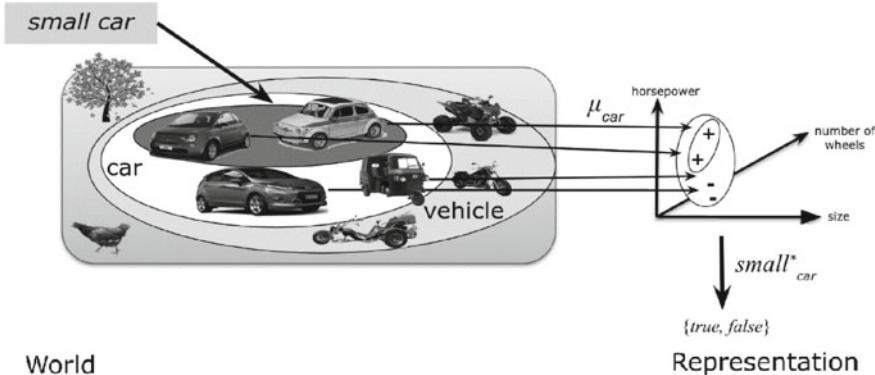
Fig. 1 A domain of vehicles and a representation featuring positive and negative exemplars of small cars

2.1 Domains and Representations

We start the formalization of our approach by introducing domains and representations. For classifiers, given the truth-value *true*, we get the extension in the attribute space by its inverse image of {*true*}, and we get its extension in the real world by applying the inverse image of the measure function. However, given a language predicate like *small* in the context of cars, its reference will in general not be completely determined by a classifier *small**$_{car}$ and by subsequently applying the inverse image of the measure function. An entity which has all the attributes of a small car may not be a small car, and an entity which is a small car may not have all the attributes we in general assign to cars. In this sense, classifiers approximate the denotation of language predicates. This approximation relation is subject to consistency constraints: If we know that x is a small car and y is similar enough to x, we expect that y is a small car, too. What should 'similar enough' mean? In our approach, we can express this in terms of the attribute space: The attribute values must be similar enough.

If the classifiers cannot discriminate between the representations (points in the attribute space) of two entities x and y, they must belong to the same concepts: If one is a small car, then the other must be a small car, too. In particular, this is the case if the representations in the attribute space are equal. Think of a situation where we measure size only with very low precision or specify color only by a few color values. If the above constraint is violated we should probably change our attribute space and/or our measure function, e.g., increase precision of measuring size and/or introduce a more fine-grained color specification.

Often we have additional structure on our attribute space, e.g., a (pre)order relation. Assume that x and y are small cars, and z is in the car domain. The number of wheels are w_x, w_y, w_z respectively; x, y, and z differ only in the number of wheels. Then, if $w_x \leq w_z \leq w_y$ we expect z to be a small car, too. If not, we again have an inconsistency in our representation. And again, we probably should change it. The mathematical

foundation of this type of inconsistency is the theory of convex closures. The formal definition of a convex closure operator cl on a set X is the following (see Korte et al. 1991):

A function $cl: \wp(X) \to \wp(X)$ is a convex closure operator iff

- it preserves the empty set $\quad cl(\{\}) = \{\}$
- it is extensive $\quad A \subseteq cl(A)$ for all $A \subseteq X$
- it is monotone $\quad A \subseteq B \to cl(A) \subseteq cl(B)$
- it is idempotent $\quad cl(cl(A)) = cl(A)$
- the anti-exchange property holds $\quad x, y \notin X, x \neq y,$
  $$x \in cl(X \cup \{y\}) \to y \notin cl(X \cup \{x\})$$

In the two-dimensional Euclidean plane, we can visualize the effect of a convex closure operator. Suppose X is $cl(\{a, b, c\})$. If x is in $cl(X \cup \{y\})$, then y cannot be in $cl(X \cup \{x\})$. The anti-exchange property ensures convexity. In a two-dimensional Euclidean plane, this means that for any two points in X the connecting line must also be in X (Fig. 2).

On a (partially) ordered set $(M, \leq)$ we can define convex closure operators in a natural way (see Fig. 3). For $A \subseteq M$ we define:

left closure: $\quad cl_{\leftarrow}(A) = \{x \in M \mid \exists\, y \in A : x \leq y\}$
right closure: $\quad cl_{\rightarrow}(A) = \{x \in M \mid \exists\, y \in A : y \leq x\}$
convex closure: $\quad cl(A) = \{x \in M \mid \exists\, y, z \in A : y \leq x \leq z\}$

To sum up: We approximate the meaning of natural language predicates by classifiers and their inverse images by means of a generalized measure function. Additionally, we request that classifiers respect some consistency constraints: (i) they should classify known examples correctly, (ii) their extension (as a subset of the attribute space) should be convex according to a suitable convex closure operator and (iii) their extensions should be open in a suitable topology. The topology and the closure operator must be compatible: Closures of open sets must be open.

Fig. 2 Convex closure and anti-exchange property in the Euclidean plane
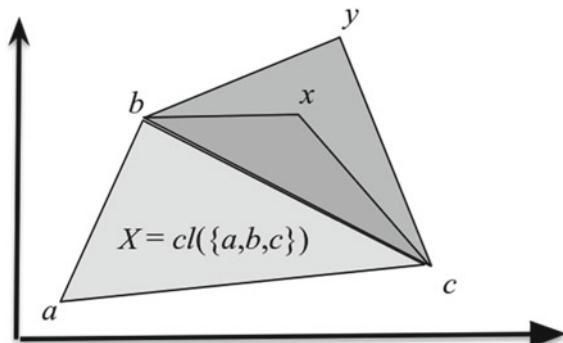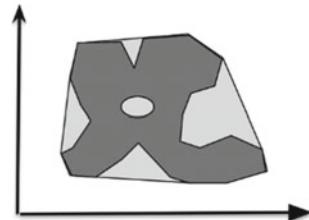
Fig. 3 A non-convex set in
the two-dimensional plane
and its convex closure

First, we need a notation to refer to the entities we are talking about by a natural language predicate like *small car*: the set of entities (in the world) for which it makes sense to ask if they have car properties, that is, entities for which the attribute dimensions for cars make sense, e.g., *number of wheels*, *horsepower*, *size*, *weight*, *color* etc. We exclude entities for which it does not make sense to ask if they have car properties, e.g., single atoms, trees, hens etc.

Next, we assume that we have clear cases: positive examples such as entities which are definitely cars, and negative examples such as entities for which the attribute dimensions of cars make sense but which are definitely not cars, e.g., motorbikes. Concepts which are related and belong to the same micro-theory are collected as predicates over the same domain. Think of different types of cars, bikes, trikes etc.

We assume that there is a universe U which includes all the entities in the world. We can start now formalizing our approach by defining a *domain* as a subset of the universe U together with a set of predicates and non-overlapping sets of positive and negative examples for each predicate.

Definition 1 *Domain*

A domain $\mathcal{D}$ is a quadruple $\langle D, \_^{+}, \_^{-}, P \rangle$ with:

- $D \subseteq U$ a set of individuals/entities (called the carrier of the domain),
- $P = \{p_1, \ldots p_n\}$ a set of identifiers of predicates over D,
- $\__D : P \longrightarrow \wp(D)$ the extension in D of a predicate[11] denoted by index $_D$[12]
- $\_^{+} : P \longrightarrow \wp(D)$ a function which assigns a set of positive examples to each predicate (for $\_^{+}(p)$ we write p^{+}),
- $\_^{-} : P \longrightarrow \wp(U)$ a function which assigns a set of negative examples to each predicate[13] (for $\_^{-}(p)$ we write p^{-}),
- $\forall p \in P : p_D^{+} \cap p_D^{-} = \emptyset$ (consistency),
- $\exists q \in P \, \forall p \in P : p_D^{+} \subseteq q_D^{+} \wedge p_D^{-} \subseteq q_D^{+} \wedge q_D^{-} \cap D = \emptyset$ (universal predicate).

---

[11] In fact, we will often use characteristic functions in place of predicates. In the structures we are interested in, there is an isomorphism between $\wp(D)$ and Ω^D. We will not restrict ourselves to a special type of logic (e.g. two-valued classical logic). We stipulate a logical system characterized by a set of truth-values Ω. $\Omega = \{true, false\}$ for classical logic, $\Omega = [0, 1]$ for fuzzy logic.

[12] We will drop the index D whenever it is clear which domain we are talking about.

[13] Positive examples must be in the domain, negative examples may be anywhere. A small mouse is a negative example for 'big elephant', but a small elephant is a more informative example.

2.2 Representations and Classifier Systems

We view the elements of D as entities to which we have only indirect access via a (generalized) measure function μ. The measure function μ constructs representations of the entities in D as points in an attribute space F, much like observables in physics. Attribute spaces are well-established representational structures.[14] They generalize vector space approaches in allowing heterogeneous dimensions equipped with value sets of different scales (nominal, ordinal, interval, proportional, partially ordered etc.), where value sets may themselves be attribute spaces with multiple dimensions.

An attribute space F is given by a set of attributes $A = \{a_1, \ldots, a_n\}$, such that for each a_i in A there is a set of possible values V_{ai} of a_i. Elements of D are mapped to points in $V_{a1} \times \cdots \times V_{an}$, the carrier of the attribute space F. Think, for example, of *number of wheels* as an attribute with $\{1, 2, 3, 4, 5, 6, \ldots\}$ as its value set, or *horsepower* as an attribute with the positive real numbers as its value set.[15]

A *representation* includes an *attribute space F*, a (generalized) *measure function* μ mapping elements of a domain into the attribute space, and a set of *classification functions p** applying to points in the attribute space. In the case of the attribute *number of wheels* the measure function μ just has to count. In the case of the attribute *horsepower* a complex measurement procedure is required to determine the value of μ. The classification functions (short *classifiers*) serve as *approximations*[16] of the predicates in P.[17] Moreover, the extensions of the classifiers will be assumed to be open and convex. This means that F comes with a convex closure operator cl and p^* must be *true* on $cl(\mu(p^+))$.[18] Using the n-dimensional Euclidean space as an example, the extensions of the classifiers must not have holes, notches or coves in the representation space F.

Definition 2 *Representation*

A representation $\mathcal{F} = \langle\langle F, cl\rangle, \mu, \_^*, \mathcal{D}\rangle$ of a domain $\mathcal{D} = \langle D, \_^+, \_^-, P\rangle$ is given by

- an attribute space F together with a closure operator cl and a compatible topology (we write F for $\langle F, cl\rangle$ if we are not interested in the closure operator cl),
- a measure function[19] $\mu: D \to F$,

---

[14]Attribute spaces are related to the classical frame approaches (Minsky 1975). Other related approaches are feature structures which are widely used in linguistic formalisms (Carpenter 1992).

[15]Note that ordinal or metric dimensions as common in degree semantics correspond to one-dimensional attribute spaces in our approach.

[16]More precisely: $p^* \circ \mu$ approximates p.

[17]For every $p \in P$ there is a $p^* \in P^*$.

[18]This includes all points in the convex closure of the images of the positive exemplars. For the concept of convexity in conceptual structures see Gärdenfors (2000). Intuitively, the convex closure of a subset X of F is the smallest convex subset of F containing X.

[19]In most cases, we do not expect to explicitly compute values of the measure function for entities in D. Almost no one will be able to compute the horse power of his car. To learn about the horse power of my car I would look-up the value in the data sheet. When you go to the doctor for a general
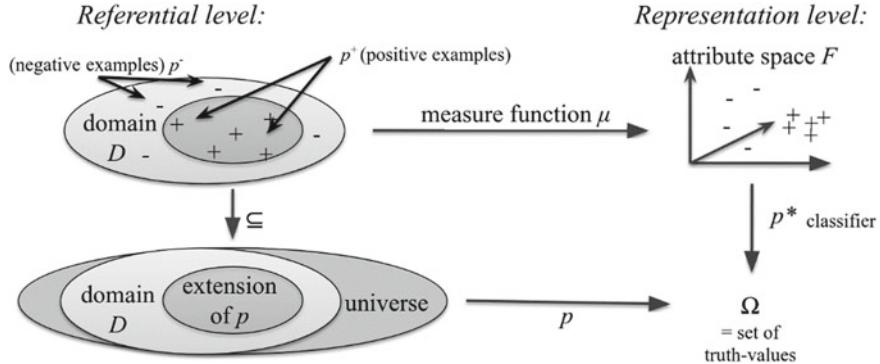
Fig. 4 Domains and representations

- a function $\_^*: P \rightarrow \Omega^F$
  (we write p^* for $\_^*(p)$ and call them classifiers).[20]

Representations are subject to three consistency constraints:

- $\forall p \in P$ the extension of p^* must be open and convex in $\langle F, cl \rangle$
- $\forall p \in P \; \forall x \in p^+: p^*(\mu(x)) = true$
- $\forall p \in P \; \forall x \in p^- \cap D: p^*(\mu(x)) = false$

From this we get $\mu(p_i^+) \cap \mu(p_i^- \cap D) = \emptyset$ (Fig. 4).

As mentioned above, attribute spaces are familiar methods of representation. What distinguishes attribute spaces from the representations proposed in this paper is the idea of classifiers on attribute spaces. On the worldy side, a domain includes a set of relevant predicates $p \in P$. On the representational side, these predicates have counterparts, namely classifiers $p^* \in P^*$. By P^* we denote the set of all basic classifiers: $P^* = \{p^* \mid p \in P\}$. These classification functions are required to be consistent with their corresponding predicates over D; more precisely, for the set of positive/negative exemplars the truth-values of the classification functions have to agree with the truth-values of the original predicates (see Definition 2).

Given a set of basic classifiers,[21] we assume the possibility to construct derived classifiers by logical operations: For the logical conjunction this is unproblematic

---

(convex sets and open sets are closed under intersection). For the logical disjunction we have to apply the convex closure operator cl to the result. For negation this is not possible. Thus we do not allow to define complex classifiers by applying negation to elementary ones.[22] We name the set of derived classifiers $\tilde{P}^*$.

Definition 3 *Classifier systems*

Given a set of basic classifiers B over an attribute space F, we define a set of classifiers $\tilde{B}$ inductively (much like a topology):

- $B \subseteq \tilde{B}$ we expect that elements of B are convex and open,
- $X \in \tilde{B}, Y \in \tilde{B} \rightarrow X \cap Y \in \tilde{B}$ intersections,
- $X \in \tilde{B}, Y \in \tilde{B} \rightarrow cl\,(X \cup Y) \in \tilde{B}$ closures of unions.

If F is (partially) ordered:

- $X \in \tilde{B} \rightarrow cl_\rightarrow(X) \in \tilde{B}$ $cl_\rightarrow(X) = \{x \in F \mid \exists\, y \in X : y \leq x\}$ right closures,
- $X \in \tilde{B} \rightarrow cl_\leftarrow(X) \in \tilde{B}$ $cl_\leftarrow(X) = \{x \in F \mid \exists\, y \in X : x \leq y\}$ left closures.

It is important to mention that in general $\tilde{B}$ is not closed under complement. This means that we do not have negation: Complements of convex sets need not be convex and complements of open sets need not be open. We start with basic classifiers $B = p^* = \{p_1{}^*, ..., p_n{}^*\}$ and get $\tilde{P}^*$ as the corresponding system of classifiers.

3   Similarity Expressions in Natural Language

In this section, a brief overview will be given of the challenges involved in the interpretation of similarity expressions. This section will not give a full description of the semantic phenomena—references will be given for details—but instead serve as a motivation for the specifics of the similarity framework presented in this paper.

3.1   Similarity Demonstratives

The need for a framework that models similarity originated from the problem of how to interpret the German demonstrative *so* ('so'/'such'). It is a genuine demonstrative

---

by using $(p_1{}^* \times \cdots \times p_n{}^*) \circ \mu$ as new measure function and Ω^n as attribute space F. Of course that is not the idea of this approach. We will try to use 'simple' measure functions and meaningful attribute dimensions.

[22] In general, complements of concepts are not necessarily themselves concepts—a non-car is not a proper concept.

expression, so we expect direct reference in the sense of Kaplan (1989). It does not, however, express identity as does, e.g., *dies/this*, and instead it refers to a set of entities which are in some sense similar to the target of the demonstration gesture (the entity the speaker points to). If the speaker points to a car while uttering "*So ein Auto hat Anna*" ('Anna has a car like this'), Anna's car is said to be, with respect to a particular set of features, indistinguishable from the car the speaker points to. This kind of demonstrative expressions is called *similarity demonstratives* in Umbach and Gust (2014), Gust and Umbach (2015), and *demonstratives of manner, quality and degree* in König and Umbach (2018).

We follow Nunberg's (1993, 2004) adaptation of the Kaplanian analysis, interpreting demonstratives as directly referential expressions, but at the same time dismissing the idea that the target of the demonstration is necessarily identical to the referent of the demonstrative. This allows for a straightforward interpretation of similarity demonstratives such that the target of the demonstration is the individual or event the speaker points to, and the referent of the demonstrative phrase is related to the target by similarity instead of identity. Similarity is then implemented by indistinguishability of points in attribute spaces (see Sect. 4). This implementation of similarity is in fact close to the idea of contextual granularization suggested in Nunberg (2004): When restricting attention to a particular set of features, it may be the case that two entities can no longer be distinguished. It is important to note, however, that this idea requires a framework that distinguishes between a referential and a representational level—you cannot speak about indistinguishability without access to what could have been distinguished.

3.2 Ad-Hoc Kinds

According to the similarity analysis, demonstratives like German *so* and English *such* create classes of similar items, e.g. similar cars. There is some evidence that in the nominal and verbal case (though not in the adjectival case) these similarity classes constitute ad-hoc kinds. In a nut-shell, *so/such* phrases can be shown to be restricted to particular features of comparison. For example, the feature *number of doors* would be perfect when comparing cars but not when comparing mugs—mugs do not have doors, so the number of doors does not qualify as a feature of comparison for mugs. But mugs as well as cars can be recently purchased and nevertheless *being recently purchased* does not qualify as a feature of comparison for neither cars nor mugs. This suggests that properties qualifying as features of comparison must not be accidental.

There is experimental evidence that features of comparison are restricted to properties which are neither accidental nor evaluative (see König and Umbach 2018; Umbach and Stolterfoht in prep.). This raises the question of how to characterize these properties, which is a prominent issue in the debate about concept formation in cognitive psychology. Only recently has this debate been connected to the topic of genericity in linguistics by Greenberg (2003) and Carlson (2010), and by the

experimental studies in Prasada and Dillingham (2006) and Prasada et al. (2013), providing evidence that there are so-called *principled connections* between kinds and properties that an entity has, because it is the kind of thing it is.

There is an alternative analysis claiming that demonstratives like German *so* and English *such* are pro-kind expressions (see Anderson and Morzycki 2015, adapting Carlson's 1980 kind-referring analysis of *such*). The final results of the two accounts are fairly close. However, unlike the pro-kind account, the similarity account not just postulates that *so/such* phrases denote kinds, but in addition shows how these kinds emerge, namely by similarity.

3.3 Equative Comparison

Another phenomenon where similarity plays a significant role is equative comparison, including non-scalar as well as scalar cases, see (3a–c).[23] In German, scalar as well as non-scalar equatives are uniformly constructed by *so … wie* where *so* is a correlative pronoun relating to the standard of comparison given in the *wie* clause:

(3) a. Anna ist so groß wie Berta   scalar/adjectival
       Anna is as tall as Berta
    b. Anna hat so ein Auto wie Berta   non-scalar/nominal
       Anna has a car like Berta's
    c. Anna tanzt so wie Berta   non-scalar/verbal
       Anna is dancing just like Berta

Given that the demonstrative *so* can in general be substituted by *wie dies* ('like this'), it suggests itself to analyze *wie* as expressing similarity as does *so*, though without a deictic component. This allows for a generalized account of equative comparison: The nominal equative in (3b) is interpreted such that Anna's car is similar to Berta's car with respect to a set of contextually given features; the verbal case in (3c) is interpreted such that the event of Anna dancing is similar to the event of Berta

---

[23]It has been argued that (3a) and (3b, c) just differ in being one-dimensional as opposed to multi-dimensional, and that even multi-dimensional comparison is scalar. There are, in fact, multi-dimensional adjectives like *healthy* that allow for comparatives: *A is more healthy than B*. Sassoon (2013) suggests to interpret comparatives of multi-dimensional adjectives by quantification over dimensions in which the compared entities exceed the standard: A is more healthy than B iff the number of dimensions in which A exceeds the standard is greater than that of B exceeding the standard (for alternatives see the subsection on gradability below).

This approach presupposes, however, that the individual dimensions are scalar, which is not generally the case, consider, e.g., color as a dimension in comparing cars or posture as a dimension in comparing dancing habits. Moreover, even though cars and dancing habits can be compared in equatives, forming comparatives is impossible. This is strong evidence that (3b,c) are genuinely non-scalar.

dancing; and the adjectival case in (3a) is interpreted such that Anna is similar to Berta with respect to their height—note that the scalar equative in (3a) does not hinge on contextually given features of comparison but instead 'carries its dimension on its sleeves'.

3.4 'Exactly' Versus 'At-Least' Reading

Scalar equatives like (3a) allow for two readings. On the *exactly* reading, Anna's height is (approximately) the same as Berta's height, while on the *at-least* reading Anna's height is greater than or equal to Berta's height. While both readings are attested in the data, standard degree semantics and the similarity analysis differ with respect to which reading is predicted to be primary. In standard degree semantics equatives are assumed to have an *at-least* interpretation as their meaning while the *exactly* reading is derived by scalar implicature. In the similarity analysis, on the other hand, equatives (scalar as well as non-scalar) are interpreted such that their meaning is symmetric, since similarity is an equivalence relation—*A ist so groß wie B* means that *A* is similar in height to *B*—thereby raising the question of how to account for the *at-least* reading.

The question of which of the *exactly* and the *at-least* reading is basic has been the topic of a continuous debate when addressing numeral expressions. According to the classic analysis by Horn (1972), sentences containing numbers assert lower boundedness and may, depending on the context, implicate upper boundedness—*Anna has three sheep* asserts that she has at least three sheep and implicates, depending on context, that she has at most three sheep. This analysis has been questioned, for example, by Kennedy (2013) who presents, among other things, scope effects that cannot be explained in the classic analysis. Surprisingly, this debate has not been extended to equative constructions, even though according to the classic analysis degree equatives assert *at-least* interpretations, as in the case of Horn's analysis of numerals: *Anna is as tall as Berta* is true if height (Anna) ≥ height (Berta) (see, e.g., Kennedy 1999).

We assume that the semantics of scalar equatives is given by similarity even in contexts requiring an *at-least* reading, and we implement this idea by exploiting the granularity encoded in our framework. Consider the example in (4). In this context, Sophie tells the truth even if she is taller than Larissa. In general, if there is a threshold given in the context, it appears irrelevant by how much it is exceeded.

(4) Sophie wants to join the police, which requires a certain minimum height. Her cousin Larissa has told their grandma that she has already been accepted by the police. That's why grandma asks Sophie whether she is as tall as Larissa. Sophie replies: Ja, ich bin so groß wie Larissa/Yes, I'm as tall as Larissa.

In the case of *at-least* readings, classifiers applying to the standard of comparison, e.g., Larissa's height in (4), are mapped to their right closure.[24] Thereby Sophie counts as similar in height to Larissa even if she is ten centimeters taller. Thus our account is "mildly ambiguous"—in particular contexts, closures involved in determining similarity are adjusted. It has to be noted, though, that this adjustment is licit only if the difference is moderate. But if, for example, Larissa is a six-year-old and Sophie is her mother, it would be absurd to assert that Sophie is as tall as Larissa (which is predicted to be true on the classical analysis of degree equatives).

For negated scalar equatives the prominent reading is asymmetrical: The sentence *Anna ist nicht so groß wie Berta/Anna is not as tall as Berta.* is preferably interpreted such that Anna is smaller than Berta. This asymmetry is not influenced by the existence of a contextual threshold and does not appear infelicitous in the case of major differences—*Larissa is not as tall as Sophie* would be acceptable even if Sophie is Larissa's mother. The preference for the asymmetric reading of negated scalar equatives can be explained by the fact that a disjunctive (symmetric) reading according to which Anna is either smaller or taller than Berta would not be convex any longer. Given that convexity plays a primary role in cognitive economy it is hardly surprising to find such effects in natural language semantics (see also Solt and Waldon 2019 on numerals under negation).

3.5 Gradability

Implementing similarity as indistinguishability (see the next section) suggests that it is a nongradable concept. This is plausible considering expressions like German *so/wie* and English *such/like*. On the other hand, the adjectives *ähnlich* and *similar* are gradable—Anna can be more similar to her father than to her mother. This points to the need for a gradable notion of similarity.

Cognitive Science models of similarity usually start out either from a notion of distance in a geometrical space (e.g. Gärdenfors 2000) or from numbers of common and distinctive features (e.g. Tversky 1977). Both approaches facilitate a straightforward definition of the comparative: In geometric models similarity increases if distances decreases, and in feature based models similarity increases if the number of common features increases and that of distinctive features decreases. However, the positive form—the predicate *similar*—would require a threshold from where on two items count as similar, which would be hard to provide in a non ad-hoc fashion.

In our system, the positive form is the primary one—two items are similar if indistinguishable with respect to a given representation (including dimensions of comparison and classifiers, see Definitions 2, 4 and 5). The comparative will be defined making use of representations of different granularity: Two items a and b are more similar than two items c and d in a representation $\mathcal{F}$ if and only if there is

---

[24]See the *quasi exactly* implementation of the *at-least* reading by right closure of classifiers in Sect. 4.

a less granular representation $\mathcal{F}'$ such that a and b are similar in $\mathcal{F}'$ while c and d are not (see Definition 8 in Sect. 5). Suppose, for example, that in representation $\mathcal{F}$ neither a and b nor c and d are similar. If there is a less granular representation $\mathcal{F}'$ such that a and b are similar while c and d can still be distinguished, then a and b must be closer in terms of properties than c and d.

Defining a comparative notion *more similar* based on the positive form *similar* is reminiscent of the vague-predicate approach suggested by Klein (1980). In contrast to the standard degree-semantic approach where degrees are compared in interpreting the comparative—*Anna is taller than Berta* is true if her degree of height exceeds that of Berta—in a Kleinian approach the comparative is modelled by varying contexts, that is, varying thresholds for the positive predicate to apply: *Anna is taller than Berta* is true if there is a context such that Anna counts as tall while Berta does not.[25] This way of interpreting the comparative is, first of all, consistent with cross-linguistic findings showing that the majority of languages express the comparative in terms of the positive. Moreover, it does not rely on the existence of a single scale of degrees.

The definition of *more similar* suggested above gives us the means to interpret the comparative form of the adjective *similar*. But beyond that it allows a Kleinian style definition of comparatives for multi-dimensional adjectives like *healthy* and *beautiful*. Comparatives of multi-dimensional adjectives are usually interpreted using degree semantics, either by counting dimensions in which the threshold is exceeded (see Sassoon 2013), or by integrating dimensions such that the result forms an order, where integration may be context-dependent and also judge-dependent (see Solt 2016).

The similarity framework puts us in the comfortable position of not having to treat all adjectives in the same way. Adjectives like *tall* and *old*, which clearly refer to a single ordinal or even metric scale, will be interpreted via a single dimension. In this case, similarity takes the role of specifying the granularity of this scale: *Anna is taller than Berta* is true if all points of the granule of Anna's height are greater than all points of the granule of Berta's height (in the case of overlapping the situation is more complex). Multi-dimensional adjectives like *healthy* and *beautiful*, on the other hand, will be interpreted by similarity to a prototype[26]: *Anna is healthy* is true if Anna's health is similar to the prototype. And *Anna is more healthy than Berta* is true if Anna's health is more similar to the prototype than Berta's health.

4 Indiscernability

In order to realize that two entities in the world are different their representations must differ in some way. This means that they must be recognizably different. In our approach this means that there are classifiers which can discriminate them. The

---

[25] Contexts have to be consistent with the order of individuals in the domain.

[26] Analogous to thresholds in a single dimension—context-dependent and maybe judge-dependent.

complementary situation is indistinguishability, which means that, on the representational level, we cannot discriminate them. In our approach, given a system of predicates P there are two reasons why we may not be able to distinguish two elements of D:

- Two elements may lead to the same value of the function μ, i.e., the same point in the attribute space. Then no classifier can discriminate between the two elements.
- The two elements disagree on μ (so we see that they are different), but they agree on all classifiers in $\tilde{P}^*$.

To account for these types of indistinguishability we borrow the term *indiscernible* from Rough Set Theory (Pawlak 1998):

Definition 4 *Indiscernible*
Given a representation $\mathcal{F} = \langle F, \mu, \_^*, \langle D, \_^+, \_^- , P \rangle \rangle$ we define:
For $x, y \in F$: $x \sim_{\mathcal{F}} y \equiv \forall q \in \tilde{P}^*$: $q(x) \longleftrightarrow q(y)$

where $\tilde{P}^*$ is the set of all derived classifiers.

According to this definition, indiscernibility is relative to the classifiers in $\tilde{P}^*$ in a representation $\mathcal{F}$. The relation of indiscernibility talks about points in F. However, the similarity relation we are interested in talks about elements of the domain D. Therefore, we have to apply the measure function before checking indiscernability. This gives us a first simple similarity relation:

Definition 5 *Similar*
$\forall x, y \in D$: $sim(x, y, \mathcal{F}) \equiv \mu(x) \sim_{\mathcal{F}} \mu(y)$

Obviously, Definition 5 defines an equivalence relation on D and we get a partition of the domain. The indiscernibility relation provides attribute spaces with a level of granularity, facilitating comparison of attribute spaces of distinct granularity which are otherwise identical. Let $[y]$ denote the equivalence class (similarity class) of y: $[y] = \{x \mid x \sim_{\mathcal{F}} y\}$. In Rough Set theory, such equivalence classes are called granules.

There is a problem with this definition of similarity: The similarity classes in the attribute space may not be convex, as the following example shows. Think of case (3a) *Anna ist so groß wie Berta* ('Anna is as tall as Berta.'). Assume that we have a dimension of height (measured in meter) in the attribute space and classifiers which specify height with some granularity depending on the measured value: A height of 1.80 is given by some value between 1.78 and 1.82, while a height of 1.81 is given by some value between 1.806 and 1.814, and so on. Therefore, we may not be able to discriminate between 1.80 and 1.815: both belong to the same granule [1.80]. Nevertheless, we can discriminate between 1.80 and 1.81 since we have a classifier [1.81] giving *true* on 1.81 and *false* on 1.80. Therefore, the granule of Berta's height ([y] in Fig. 5, which is equal to [1.80]) may be not convex because [1.81] forms a hole. This results in the following situation: If Berta's height is 1.80, then Anna's height may be 1.80 or 1.815 but not 1.81 in order for the sentence to be true (as demonstrated in Fig. 5). This is counterintuitive.
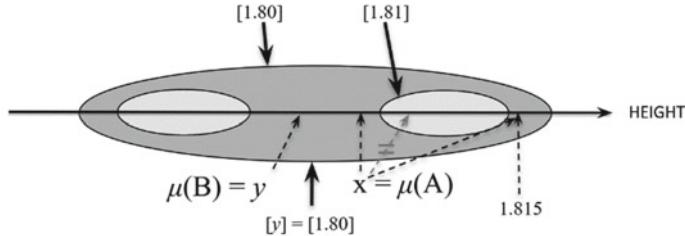
Fig. 5 Granules with holes

We can solve this problem by introducing a new parameter in the definition of the similarity relation: *similarity relative to a point of reference*. This point of reference determines the granules to be selected.

Definition 6 *Similarity relative to a point of reference*

Given a representation $\mathcal{F} = \langle \langle F, cl \rangle, \mu, \_^*, \langle D, \_^+, \_^-, P \rangle \rangle$, we can define a similarity relation relative to a point of reference r in two different ways:

$$\forall x, y \in F: x \sim_{\mathcal{F}_r} y$$

(a) iff $\forall q \in \tilde{P}^* : q(r) \to q(x) \wedge q(y)$
(b) iff $\forall q \in \tilde{P}^* : q(r) \to (q(x) \leftrightarrow q(y))$

Definition 6a means that principal filters[27] of x and y in $\tilde{P}^*$ contain the principal filter of r. In contrast, Definition 6b means that elements of the principal filter of r in $\tilde{P}^*$ cannot discriminate between x and y. It is easy to see that (a) $\Rightarrow$ (b), but not (b) $\Rightarrow$ (a).

For an intuitive insight into the functionality of this type of similarity relation, have a look at the Venn diagrams in Fig. 6 and at Table 1:
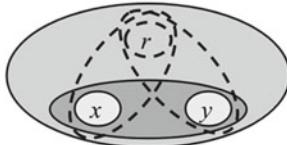
Assume that there are four classifiers in $\tilde{P}^*$: *small**, *big**, *normal** (concerning size), and *heavy** (concerning weight). Table 1 shows some possible classifications of x, y, and r. These possibilities correspond to the dashed sets in Fig. 6. The last two columns show the truth-values of the two similarity relations (a) and (b) in Definition 6 for the different cases. All the other cases can be handled by symmetry; only *heavy** varies. The interesting case is line (2) since the two similarity relations differ: If y is small but r and x are not, and x is big but r and y are not, and x and y are normal but r is not, and r is heavy but x and y are not, then similarity of x and y with respect to the reference point r is *true* according to Definition 6b but *false* according to Definition 6a. Intuitively, if the properties of the reference point r differ substantially from the properties of x and y then Definition 6a gives *false* while 6b gives *true*. We consider Definition 6a more plausible than 6b.

For given $\mathcal{F}$ and r the relation $x \sim_{\mathcal{F}_r} y$ is a (kind of local) equivalence relation. If we switch the reference r, the classes will obviously change. If we choose one of the arguments as point of reference, we get an asymmetric similarity relation: In general $x \sim_{\mathcal{F}_y} y$ will be different from $y \sim_{\mathcal{F}_x} x$ because the point of reference changes.

---
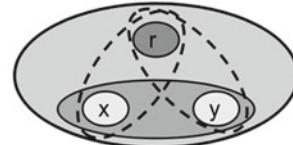
[27]The principal filter of x is $\{q \in \tilde{P}^* \mid q(x)\}$.

Table 1 Similarity of two points x and y in the attribute space with respect to a reference point r depending on the possible extensions of the predicates *small**, *big**, *normal**, and *heavy**. The cell [*small**, (1)], for example, indicates that *small**(r) is false, *small**(x) is false and *small**(y) is true. The cell [*heavy**, (1)] indicates that *heavy**(r) is false, *heavy**(x) is either true or false, and *heavy**(y) is also either true or false

	small*	*big**	*normal**	*heavy**	$x \sim_{\mathcal{F}r} y$	
	r x y	r x y	r x y	r x y	(a)	(b)
(1)	− − +	− + −	− + +	− * *	true	true
(2)	− − +	− + −	− + +	+ − −	false	true
(3)	− − +	− + −	− + +	+ + −	false	false
(4)	− − +	− + −	− + +	+ − +	false	false
(5)	− − +	− + −	− + +	+ + +	true	true

If $x \sim_{\mathcal{F},r} y$ holds, classifiers of the dashed sets cannot occur in $\tilde{P}*$

(a) $\forall q \in \tilde{P}* : q(r) \rightarrow q(x) \wedge q(y)$

$\neg\neg \forall q \in \tilde{P}* : q(r) \rightarrow q(x) \wedge q(y)$

$\neg \exists q \in \tilde{P}* : q(r) \wedge (\neg q(x) \vee \neg q(y))$

(b) $\forall q \in \tilde{P}* : q(r) \rightarrow (q(x) \leftrightarrow q(y))$

$\neg\neg \forall q \in \tilde{P}* : q(r) \rightarrow (q(x) \leftrightarrow q(y))$

$\neg \exists q \in \tilde{P}* : q(r) \wedge \neg(q(x) \leftrightarrow q(y))$

Fig. 6 If dashed sets occur in $\tilde{P}*$, x and y cannot be similar

Definition 7 *Similarity classes*

For given $\mathcal{F}$ and r we define the similarity class of r as

(a) $[r]_{\mathcal{F}} = \{x \mid \forall q \in \tilde{P}*: q(r) \rightarrow q(x)\}$.

For $[r]_{\mathcal{F}}$ we borrow the term granule from Rough Set theory. Again we can use the inverse image of the measure function to define similarity relations on the domain. For $a, b \in D$ we define two different similarity relations. The one in (b) makes use of a point of reference r that is independent of either a or b, whereas in (c) the point of reference is identical to the second argument:

(b) $sim_r (a, b, \mathcal{F})$ iff $\mu(a) \sim_{\mathcal{F}r} \mu(b)$ (+transitive, + symmetric, −reflexive)

(c) $sim' (a, b, \mathcal{F})$ iff $\mu(a) \sim_{\mathcal{F}b} \mu(b)$ (−transitive, −symmetric, +reflexive) [28]

If we again look at our example (3a) *Anna ist so groß wie Berta* ('Anna is as tall as Berta.') we see that the granules depend on the point of reference r (Fig. 7). If we use *sim'* from Definition (7c), there are two possible situations. In the first situation, we get the information that the height of Berta is 1.80. Since Berta provides the reference point (Definition 7c) the relevant granule is [1.80]. The height of Anna can

[28] *sim'* uses the second argument as point of reference.
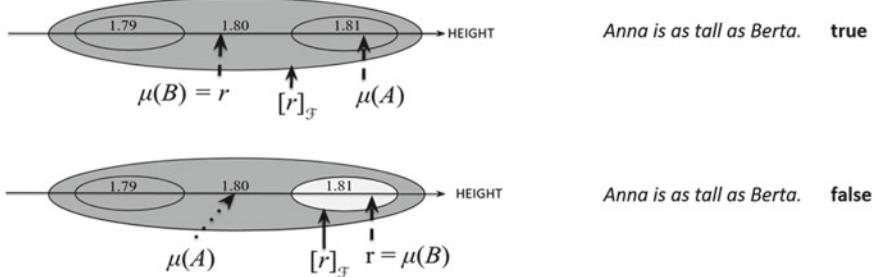
Fig. 7 The effect of holes

be an arbitrary value in this granule to make the statement true. It maybe 1.80 or 1.81—we simply cannot discriminate between both cases because the granule [1.80] is convex (no holes). In the second situation, we get the information that the height of Berta is 1.81. Now the relevant granule is [1.81] and not [1.80] even though 1.81 may be an element of [1.80]. The height of Anna is restricted to the relevant granule: 1.80 is not a possible value any longer, it falsifies the statement. Although it seems that there is a hole in [1.80] in the second case, in both cases, the relevant granule is convex.

4.1 (A)symmetry of Similarity

The notion of *similarity relative to a reference point* is reminiscent of the question of whether the predicate *similar* is symmetrical addressed by Tversky (1977) and also Gleitman et al. (1996).

Tversky's seminal paper on feature-based similarity starts with empirical observations indicating problems of the then predominant geometric notion of similarity and the basic axioms of metric distance[29]: (i) minimality is problematic in view of results concerning the identification probability for identical stimuli, (ii) symmetry is apparently false—the judged similarity of North Korea to Red China exceeds the judged similarity of Red China to North Korea—and (iii) triangle inequality is hardly compelling—Jamaica is similar to Cuba (geographical proximity) and Cuba is similar to Russia (political affinity) but Jamaica and Russia are not similar at all.

However, a closer look reveals that these findings are not generally valid. Before dismissing transitivity of the similarity relation on the basis of the Jamaica/Cuba/Russia example, one should consider the role of switching features within the two comparison steps.[30] And before dismissing symmetry, which is

---

[29] A metric distance function δ has to comply with (i) minimality: $\delta(a, b) \geq \delta(a, a) = 0$, (ii) symmetry: $\delta(a, b) = \delta(b, a)$ and (iii) triangle inequality: $\delta(a, b) + \delta(b, c) \geq \delta(a, c)$.

[30] sim' (Definition 7c) is in fact intransitive due to using the second argument as point of reference.

frequently done in the Cognitive Science literature, one should consider the study in Gleitman et al. (1996) and, first of all, Tversky's original study.

In Tversky's study, the linguistic presentation was directional (*North Korea is similar to Red China*), and he himself argues that the asymmetry finding hinges on the directional way of presentation. If the task is to assess the degree to which *A* is similar to *B*, then features of *A* may weigh more heavily than those of *B*.[31,32] But if the task is to assess the degree to which *A* and *B* are similar to each other, weights are expected to be equal and similarity judgements are symmetric. In Gleitman et al. (1996) the influence of directional vs. nondirectional presentation is experimentally examined for a number of predicates that are intuitively thought to be symmetrical including *similar*, *equal* and *identical*. The authors find that the way of presentation is decisive for the (a)symmetry in the interpretation of these predicates, even if the nouns they are combined with are nonsense nouns.

Tversky as well as Gleitman et al. attribute the asymmetry effects triggered by directional presentation to the difference between Figure and Ground. The same idea is found in our second definition of relative similarity (Definition 7c), where the second argument takes the role of the Ground in determining the relevant granule.

4.2 'Exactly' Reading Versus 'At-Least' Reading

As shown in Sect. 3, scalar equatives may have two readings: an *exactly* reading and an *at-least* reading—*Anna is as tall as Berta* may be interpreted such that Anna's height is the same as Berta's height or such that Anna's height exceeds Berta's height. We assume that the semantics of scalar equatives is uniformly given by similarity even in contexts requiring an *at-least* reading, and we implement this idea by exploiting the granularity provided by closures on classifier systems.

The *exactly* reading of equatives is accounted for by the granules defined by the available classifiers and the reference point $\mu(Berta)$. $\mu(Anna)$ must be in the granule of $\mu(Berta)$. To account for the *at-least* reading we need a transformation of classifiers such that all degrees above a certain point x count as similar.[33] Formally, we define a mapping from the classifier set $\tilde{P}^*$ to a subset $\tilde{P}^*_x$ such that every p^* in $\tilde{P}^*$ that classifies a member of $cl_\rightarrow([r]_\mathcal{F})$ as true is mapped to its right closure while the others stay unchanged. Figure 8 shows such a mapping: All classifiers left

---

[31] In Tversky's contrast model a function *S* takes weighted sums of the feature sets *A* and *B* of objects *a* and *b* to an interval scale such that $sim(a, b) \leq sim(c, d)$ iff $S(a, b) \leq S(c, d)$, where $S(a, b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$, α, β, θ denote weighting functions and *f* denotes a nonnegative scale.

[32] There is also the issue of which features are activated in the first place. In a directional presentation the subject will determine which features are relevant in comparison.

[33] If we have a simple interval scale, we can model the *at-least* reading directly by the order of the attribute values. If we want to model granularity in addition, it becomes more complex since granules may overlap. If the scale is weaker or multiple dimensions are involved, comparison becomes even more complex. Our approach provides a uniform framework for all these cases.
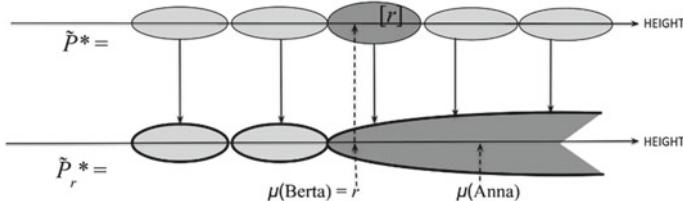
Fig. 8 *Quasi-exactly* implementation, one dimension

of [r] stay unchanged, while all classifiers to the right of [r] will be mapped to [r]. If the classifier extensions overlap, the situation may be quite complex. The right closure of [r] handles the general case. This procedure makes it possible to derive the *at-least* reading from the *exactly* reading by solely adapting classifiers. We call it a *quasi-exactly* implementation of the *at-least* reading:

Quasi-exactly **implementation of the** *at-least* **reading** by right closure of classifiers:
  $\tilde{P}_r^* = \{p_r^* \mid \text{for } p^* \in \tilde{P}^* \text{ if } p^* \cap cl_\rightarrow([r]_\mathcal{F}) \neq \emptyset \text{ then } p_r^* = cl_\rightarrow(cl(p^* \cup [r]_\mathcal{F}))$
else $p_r^* = p^*\}$.

Although we get an *at-least* reading, the result still defines an equivalence class[34]: If we select a granule by a point of reference, every element in the granule is equivalent to every other element in the granule. This approach can handle multi-dimensional cases, too. Assume that we are talking about the size of tables represented by dimensions *length* and *width*, and we use the classical convex closure of the Euclidean two-dimensional space. For non-overlapping classifiers the following two situations may occur (Fig. 9a, b). If the extension of a classifier p^* is outside $cl_\rightarrow([r])$, then p^* stays unchanged. If it is inside, then p^* will be mapped to $cl_\rightarrow([r])$, analogous to the one-dimensional case. The general case with overlapping classifiers is again covered by the formulas in Fig. 9a, b.

It is essential in our approach that the *exactly* interpretation is the primary one and is specified by the granularity given by the (contextually determined) classifier system $\tilde{P}^*$. The *at-least* interpretation is derived by applying a transformation to the classifier system $\tilde{P}^*$ depending on the reference element r.

5 Granularity of Representations and Gradability of Similarity

As stated in Sect. 3, granularity of representations provides a notion of *more similar* serving in the interpretation of the comparative form of the adjective *similar*. More importantly, the notion of *more similar* is exploited in the interpretation of multi-dimensional adjectives in general—positive as well as comparative forms. *Anna is*

---

[34]Since we have to select the granule first, it is a kind of 'local' equivalence class.
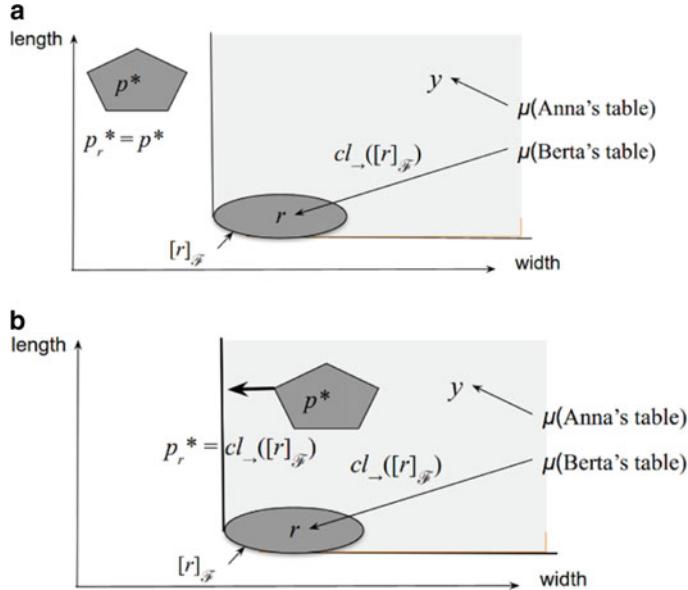
Fig. 9 a *Quasi-exactly* interpretation, two dimensions, $p^* \cap cl_\rightarrow([r]_\mathcal{F}) = \emptyset$. **b** *Quasi-exactly* interpretation, two dimensions, $p^* \cap cl_\rightarrow([r]_\mathcal{F}) \neq \emptyset$

healthy is true if Anna's health is similar to a (contextually determined) *healthy* prototype. *Anna is more healthy than Berta* is true if Anna's health is more similar to the prototype than Berta's health.

The core of the formalism are sets of *representations* equipped with a preorder structure (transitive, reflexive, but maybe not antisymmetric). This preorder implements a concept of granularity and granularity change. It will be used to construct a predicate *more_similar* based on a similarity relation defined by indiscernibility. For two representations $\mathcal{F}$ and $\mathcal{F}'$ we can ask whether one is more fine-grained than the other, that is, whether there are entities that can be distinguished in one representation but not in the other. Distinguishability is the opposite of indiscernibility and depends on the attribute spaces and the available classifiers. Therefore, these parameters determine the granularity of representations. We will introduce a reflexive and transitive relation on representations (a preorder), which relates granularity levels.

Definition 8 *Granularity of representations*
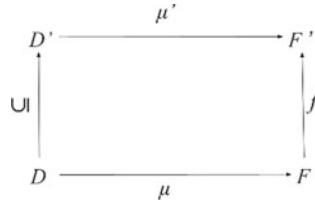
Given two representations
$$\mathcal{F} = \langle F, \mu, \_^*, \mathcal{D} \rangle \qquad \text{with } \mathcal{D} = \langle D, \_^+, \_^-, P \rangle$$
$$\mathcal{F}' = \langle F', \mu', \_^{*'}, \mathcal{D}' \rangle \quad \text{with } \mathcal{D}' = \langle D', \_^{+'}, \_^{-'}, P' \rangle$$

we define:

$\mathcal{F}'$ is *at least as coarse* as $\mathcal{F}$, $\mathcal{F}' \geq \mathcal{F}$ iff there is a function f such that

(a)   the following diagram commutes:

(b) $\forall x, y \in F: x \sim_{\mathcal{F}} y \rightarrow f(x) \sim_{\mathcal{F}'} f(y).$

This definition states that what is indiscernible in the finer representation cannot be discriminated in the coarser representation. The strict version $\mathcal{F}'$ *is coarser than* $\mathcal{F}$, $\mathcal{F}' > \mathcal{F}$, can be defined by the non-strict one:

$\mathcal{F}' > \mathcal{F}$ iff $\mathcal{F}' \geq \mathcal{F}$ and not $\mathcal{F} \geq \mathcal{F}'$

What we need now is a specification of a relevant set of representations $\mathcal{H}$. The *coarser* relation then turns $\mathcal{H}$ into a preorder. We call such a structure a *hierarchy of representations.* What is missing to get a partial order from a preorder is the anti-symmetry axiom: from $\mathcal{F} \geq \mathcal{F}'$ and $\mathcal{F}' \geq \mathcal{F}$ we cannot conclude that $\mathcal{F} = \mathcal{F}'$. We may have different possibilities to get the same structure of granules. These hierarchies are related to the concept of context (van Rooij 2011).

Definition 9 *Hierarchy of representations*

A hierarchy $\mathcal{H}$ is a set of representations such that for any two elements
$\mathcal{F}_{1/2} = \langle\langle F_{1/2}, cl_{1/2}\rangle, \mu_{1/2}, \_^*_{1/2}, \langle D_{1/2}, \_^+_{1/2}, \_^-_{1/2}, P_{1/2}\rangle\rangle \in \mathcal{H}$

we postulate the following constraints[35]:
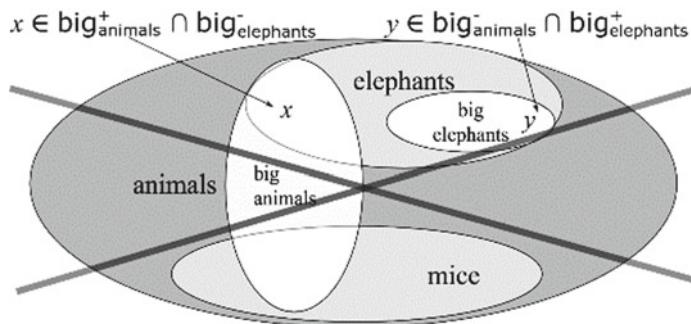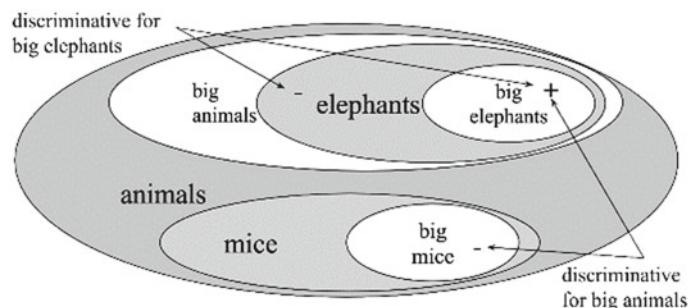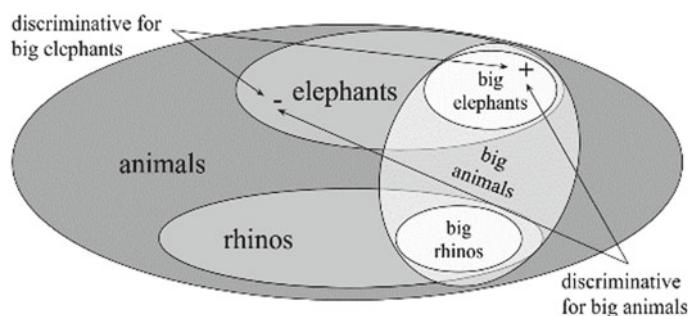
- consistency: $\forall p \in P_1 \cap P_2: (p^{+1} \times p^{-1}) \cap (p^{-2} \times p^{+2}) = \emptyset$
  Elements of p^+ and p^- cannot change roles in different domains.
- discriminative power: $\forall p \in P_1 \cap P_2: (p^{+1} \times p^{-1}) \cap (D_2 \times D_2) \neq \emptyset \rightarrow p^{-2} \times p^{+2} \neq \emptyset$
  If a domain contains a discriminating pair of another domain for a shared predicate identifier, it must itself contain a discriminating pair.[36]
- connectedness:
  $\exists \mathcal{F} = \langle\langle F, cl\rangle, \mu, \_^*, \langle D, \_^+, \_^-, P\rangle\rangle \in \mathcal{H}: D_1 \subseteq D \wedge D_2 \subseteq D \wedge P_1 \subseteq P \wedge P_2 \subseteq P$
  and there are continues closure preserving functions $f_{1/2}: F \rightarrow F_{1/2}$ with $\mu_{1/2} = f_{1/2} \circ \mu.$

For any two domains there is an enclosing domain.
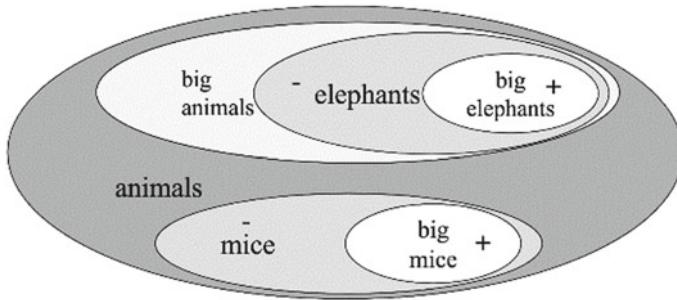
These constraints can be visualized by the Venn diagrams in Fig. 10–13:

---

[35](a) and (b) are adaptations of the context constraints in (van Rooij 2011: Definition 1).

[36]If we have big and small elephants and view them as animals, then there should be big and small animals, too. Either there are small animals like mice or, if all animals have the size of elephants, then small elephants must be small animals, too. See the Venn-diagram in Figs. 11 and 12.

Fig. 10 Consistency

Fig. 11 Discriminative power, 1

Fig. 12 Discriminative power, 2

(a)   the **consistency** constraint rules out cases like this: if y is a big elephant and x
       is a small (not big) one, then x cannot be a big animal if y is a small (not big)
       one (Fig. 10).

Fig. 13 Connectedness

(b) **discriminative** power: If there are big and small elephants there must be big and small animals, too, because animals are different in size: We have big and small elephants which are animals.

    b1.   If we collect elephants and mice in one animal domain, then a mouse (big or not) is a negative example for big animals. Thus we have a discriminating pair for big animals (Fig. 11).

    b2.   If we collect only big animals in one animal domain, say elephants, hippos, and rhinoceroses, then any discriminating pair for these species is also discriminative for big animals (Fig. 12).

(c) **connectedness**: For any two domains there must be a super domain containing both (upward directed) (Fig. 13).

In the remainder of this section we assume that there is a contextually given hierarchy of representations $\mathcal{H}$. Our approach is non-constructive in the following aspect: We do not construct representations and hierarchies, but instead have systems of constraints which hierarchies must obey. The instantiations must be given by, e.g., the situation of the utterance.

We will now demonstrate how to define a general relation $more\_sim(a, b, c, d, \mathcal{F})$ based on our similarity relation sim and the preorder on representations. The relation $more\_sim(a, b, c, d, \mathcal{F})$ is intended to be true if a is more similar to b than c is to d with respect to a representation $\mathcal{F}$.

Definition 10 *More similar*

Given a hierarchy $\mathcal{H}$, a similarity relation[37] sim, and a representation $\mathcal{F} \in \mathcal{H}$, we define
$more\_sim(a, b, c, d, \mathcal{F})$ iff

(a) $\exists \, \mathcal{F}' \in \mathcal{H}: \mathcal{F}' \geq \mathcal{F} \land sim(a, b, \mathcal{F}') \land \neg \, sim(c, d, \mathcal{F}')$

(b) $\forall \, \mathcal{F}' \in \mathcal{H}: \mathcal{F}' \geq \mathcal{F} \rightarrow (sim(c, d, \mathcal{F}') \rightarrow sim(a, b, \mathcal{F}'))$

---

[37] We discussed different similarity relations (see Sect. 4). In this definition, we can use any of these.

The widely used version $more\_sim(a, b, c, \mathcal{F})$ in the sense that a is more similar to b than c is similar to b can be defined straightforwardly by:

$more\_sim(a, b, c, \mathcal{F}) \equiv more\_sim(a, b, c, b, \mathcal{F})$

If a is more similar to b than c to d in a given representation $\mathcal{F}$ it must be possible to discriminate between c and d. Otherwise, because c and d are maximal similar, a and b cannot be more similar than c and d. If we can discriminate between c and d in $\mathcal{F}'$ then we can discriminate between c and d in every finer representation but maybe not in every coarser one. If we can find a representation $\mathcal{F}'$ (maybe coarser than $\mathcal{F}$), such that we can discriminate between c and d but not between a and b (Definition 10a), we are almost done. It remains to exclude contradictions, that is, representations in which we can discriminate between a and b but not between c and d (this is excluded by Definition 10b).

The diagrams in Fig. 14 and 15 show example hierarchies of representations talking about color and size of objects (each circle stands for a representation). We start with Fig. 14.

Representations which are higher in the hierarchy are coarser than lower ones. On the left branch we introduce a dimension *color* and a classifier system based on {*yellow**, *light-blue**, *blue**} which can classify colors by convex subsets of a (three-dimensional) color space. On the right branch, we introduce a dimension *size* with a corresponding classifier system {*small**, *big**, *huge**}. The bottom representation integrates the left branch and the right branch (Definition 9 connectedness). Again, the *size* dimension need not to be a simple proportional scale. It can itself be a three-dimensional vector space with sub-dimensions *length*, *width*, and *height*.

According to the Definition 10a, the *more_sim* relation will be inherited from top to bottom along the coarser relation. In the circles, we see the extensions of the corresponding $\tilde{P}^*$ elements. Next to the circles we see the statements about *more_sim* which are true in these representations. These statements depend not only on the representation they are attached to, but on the whole upper structure (the filter) of the representation. If we look at the circle at the bottom $\mathcal{F}_{c+s}$, we see that we inherit two statements, both from the left branch:

$more\_sim(y, z, x, \mathcal{F}_{c+s})$ and $more\_sim(z, y, x, \mathcal{F}_{c+s})$.

From the right branch, we inherit nothing because the classifier system is too weak. Representations may inherit inconsistent information from different paths which rule out some of the statements (by Definition 10b). We can see this when we add more powerful classifiers to the right branch, see Fig. 15.

The two heavily bordered circles ($\mathcal{F}_L$ and $\mathcal{F}_R$) are alternatives which have different effects on the more fine-grained representations (below). The representation $\mathcal{F}_s$ (circle below $\mathcal{F}_L$ and $\mathcal{F}_R$) inherits *more_sim* statements though some are ruled out by the consistency constraint (Definition 10b). In the bottom circle $\mathcal{F}_{c+s}$ all statements are ruled out by the consistency constraints if both $\mathcal{F}_L$ and $\mathcal{F}_R$ are present in $\mathcal{H}$. In $\mathcal{F}_{c+s}$, $more\_sim(z, y, x, \mathcal{F}_{c+s})$ would be true (z is more similar to y than x is) according to color because of $\mathcal{F}_c''$ and Definition 10a. In this case, we cannot discriminate between z and y, but we can discriminate between x and y. According to the existential quantifier in Definition 10a, this is propagated downwards. On the
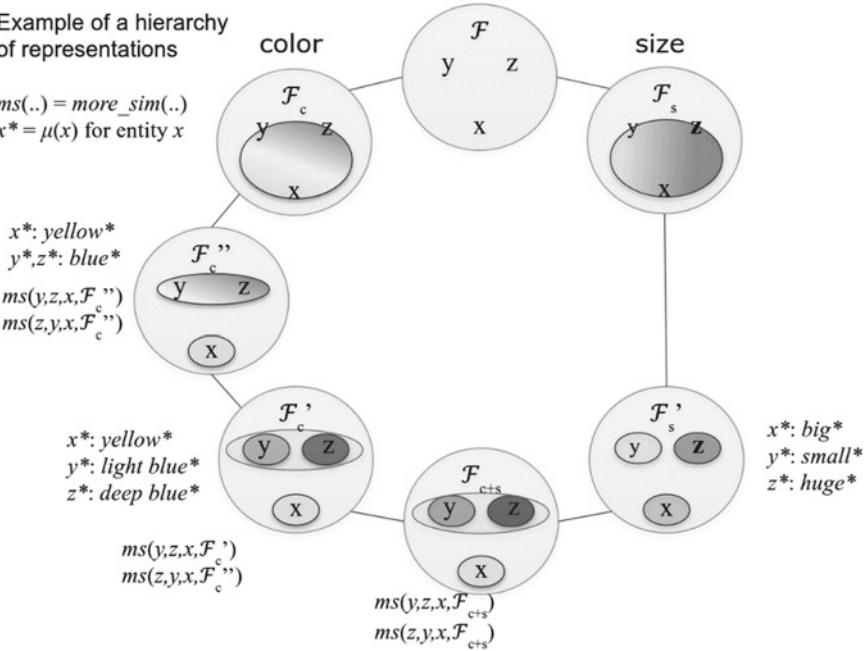
Fig. 14 Hierarchy of representations, Example 1
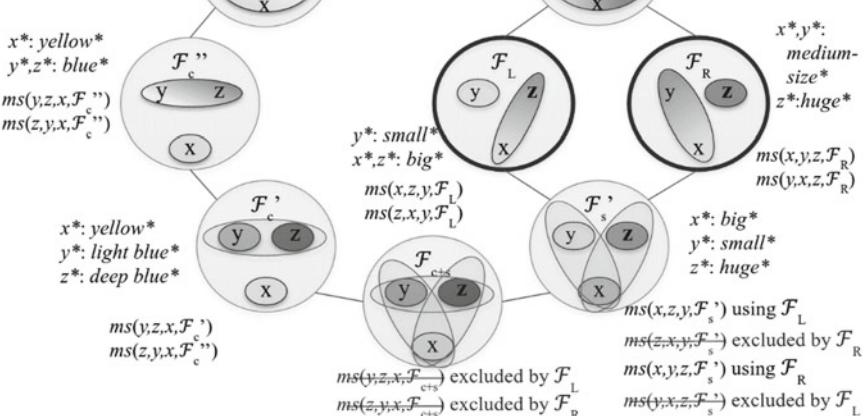
Fig. 15 Hierarchy of representations, Example 2

other side, in $\mathcal{F}_R$ we cannot discriminate between *x* and *y*. According to the Definition 10b and the universal quantification we should not be able to discriminate between *z* and *y* in this representation, but we are. Therefore, we get a contradiction.

Since in a natural language utterance the hierarchy of representations is not explicitly expressed, we can interpret the meaning of an utterance like *A is more similar to B than C* only as constraint on the relevant hierarchy of representations.

6 Conclusion

We presented a framework introducing a non-metric and qualitative concept of similarity suitable for the interpretation of similarity in natural language.

The basic idea is to "measure" properties of individuals with the help of multi-dimensional attribute spaces representing relevant features of comparison (thus generalizing the idea of degree semantics). In our framework, attribute spaces are complemented by classifiers which are predicates on points in attribute spaces approximating domain predicates; this is what we define as a representation. Individuals count as similar with respect to a particular representation if their values are indistinguishable.

In our framework, the granularity of the similarity relation may vary due to different dimensions of comparison and classifier systems. This leads to sets of representations forming hierarchies of different granularity levels, where the order on representations facilitates a Kleinian style notion of *more similar*.

This system provides a powerful and flexible tool to capture the meaning of natural language similarity expressions and account for the role of similarity in ad-hoc kind formation as well as equative comparison. Future work will explore its capacity in, e.g., multi-dimensional comparison of adjectival, nominal and verbal properties. The general idea of our approach is to reconstruct comparison in natural language in a qualitative way, with the help of different levels of granularity imposed by constraints on systems of classifiers.

References

Anderson, C., & Morzycki, M. (2015). Degrees as kinds. *NLLT, 33*(3).

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11*(3), 211–222.

Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In E. Kittay & A. Lehrer (Eds.), *Frames, fields, and contrasts.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Carlson, G. (1980). *Reference to kinds in English.* Garland.

Carlson, G. (2010). Generics and concepts. In F. J. Pelletier (Ed.), *Kinds, things and stuff* (pp. 16–36). Oxford: Oxford University Press.

Carpenter, B. (1992). *The logic of typed feature structures.* Cambridge Tracts in Theoretical Computer Science, Cambridge University Press.

Gärdenfors, P. (2000). *Conceptual spaces.* MIT Press.

Gleitman, L., Gleitman, H., Miller, C., & Ostrin, R. (1996). Similar, and similar concepts. *Cognition, 58*, 321–376.

Greenberg, Y. (2003). *Manifestations of genericity.* Routledge.

Gust, H., & Umbach, C. (2015). Making use of similarity in referential semantics. In H. Christiansen, I. Stojanovic & G. Papadopoulos (Eds.), *Proceedings of the 9th Conference on Modeling and Using Context,* Context 2015. LNCS. Springer.

Horn, L. (1992). The said and the unsaid. In B. Chris, & D. David (Eds.), *Proceedings of Semantics and Linguistic Theory 2* (pp. 163–192). OSU Working Papers in Linguistics.

Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry & H. Wittstein (Eds.), *Themes from Kaplan* (pp. 481–563). Oxford University Press.

Kennedy, C. (1999). *Projecting the adjective.* Garland Press.

Kennedy, C. (2013). A scalar semantics for scalar readings of number words. In I. Caponigro, & C. Cecchetto (Eds.), *From Grammar to Meaning: The Spontaneous Logicality of Language.* Cambridge University Press.

Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics & Philosophy, 4*(1), 1–45.

König, E., & Umbach, C. (2018). Demonstratives of manner, of quality and of degree: A neglected subclass. In M. Coniglio, E. Schlachter & T. Veenstra (Eds.), *Atypical demonstratives: Syntax, semantics and typology.* Berlin, de Gruyter Mouton.

Korte, B., Lovász, L., & Schrader, R. (1991). *Greedoids.* Springer.

Minsky, M. L. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision.* New York.

Nunberg, G. (1993). Indexicality and deixis. *Linguistics and Philosophy, 16*(1), 1–43.

Nunberg, G. (2004). Indexical descriptions and descriptive indexicals. In M. Reimer & A. Bezuidenhout (Eds.), *Descriptions and beyond.* Oxford University Press.

Pawlak, Z. (1998). Granularity of knowledge, indiscernibility and rough sets. In *Proceedings of the IEEE International Conference on Fuzzy Systems* (pp. 106–110).

Prasada, S., & Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. *Cognition, 99*, 73–112.

Prasada, S., Khemlani, S., Leslie, S.-J., & Glucksberg, S. (2013). Conceptual distinctions amongst generics. *Cognition, 126*, 405–422.

van Rooij, R. (2011). Implicit versus explicit comparatives. In P. Égré & N. Klinedinst (Eds.), *Vagueness and language use* (pp. 51–72). Palgrave Studies in Pragmatics, Language and Cognition. Palgrave Macmillan, London.

Sassoon, G. (2013). A typology of multidimensional adjectives. *Journal of Semantics, 30*, 335–380.

Solt, S. (2016). Multidimensionality, subjectivity and scales: experimental evidence. In E. Castroviejo, L. McNally & G. Sassoon (Eds.), *Gradability, scale structure and vagueness: Experimental perspectives.* Springer.

Solt, S., & Waldon, B. (2019). Numerals under negation: Empirical findings. *Glossa, 4*(1), 113.

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

Umbach, C., & Gust, H. (2014). Similarity demonstratives. *Lingua, 149*, 74–93.

Umbach, C., & Gust, H. (in print). Grading similarity. In S. Löbner, T. Gamerschlag, T. Kalenscher, M. Schrenk & H. Zeevat (Eds.), *Concepts, frames and cascades in semantics, cognition and ontology.* Springer.

Umbach, C., & Stolterfoht, B. (in prep.). Ad-hoc kind formation by similarity.

Numerical Concepts in Context

Paola Gega, Mingya Liu, and Lucas Bechberger

1 Introduction

Numerical concepts are an integral part of everyday conversation and communication. While mathematicians assign a precise interpretation to a natural number, e.g., 5 being exactly 5, the use and understanding of numerical expressions in natural language have a high variability. Broadly speaking, scientists use numbers more precisely when they discuss their research results (for example, 0.051 and 0.049 make a big difference in term of statistical significance) than street vendors at a flea market of Berlin (e.g., 51 or 49 cents for a broken antique glass are probably equally good results). In addition to broad context, narrower context such as questions under discussion (QUDs, Roberts 1996) or decision problems can influence the interpretation of numerical expressions as well: If a waiter asks "How many beers would you like to order?", we mean exactly 10 when we say 10, no more no less. If a student is eligible for taking the exam with 2 assigned tasks, s/he is eligible with 2 assigned tasks—2 means at least 2. In contrast, if a student can pass the exam with 10 mistakes, 10 means at most 10. Furthermore, the interpretation of numerical expressions can also be subject to individual and developmental factors (e.g., Musolino 2004). In

P. Gega (✉)
Institute of Philosophy, University of Bochum, Bochum, Germany
e-mail: paola.gega@rub.de

M. Liu (✉)
Department of English and American Studies, Humboldt University of Berlin, Berlin, Germany
e-mail: mingya.liu@hu-berlin.de

L. Bechberger
Institute of Cognitive Science, Osnabrück University, Osnabrück, Germany
e-mail: lucas.bechberger@uni-osnabrueck.de

this paper, we will focus on the interpretive variability of numerical expressions in narrow linguistic contexts, namely, the nature of a number itself, and its co-occurring expressions.

Among others, the interpretation of numerical expressions depends on the perceived "roundness": Round numbers (e.g., 50) can have both an imprecise or precise interpretation, whereas non-round numbers (e.g., 47) tend to have a precise interpretation. For example, Krifka (2002, 2007, 2009) proposes a "RNRI" (round numbers round interpretation) principle: "Round number words tend to have a round interpretation in measuring contexts". Supporting evidence comes from the highly frequent use of round numbers in, among others, newspapers or street/distance signs, even though statistically speaking, it is very unlikely that the results of measurements are round more frequently than they are not (given sensitive instruments). In (1), taken from the Leipzig Wortschatz Corpus (Goldhahn et al. 2012), it is intuitive to assume that all the numerical expressions have an imprecise interpretation.

(1)   a. **Forty thousand people** in the state remained without water, and **26,000 people** were without electricity, she said, warning once again that people should stay inside.
       b. Gibraltar Airport - Located just **500 meters** from the city center, Gibraltar's airport landing strip shares space with one of the island's main roads.

Another piece of evidence is shown in the contrast between (2a) and (2b). Whereas (2a) is acceptable to characterize situations where John made 49 cupcakes, the use of (2b) is degraded in the same contexts. This shows that in contrast to round numbers, non-round numbers have a precise interpretation.

(2)   a. John made **50 cupcakes**.
       b. John made **48 cupcakes**.

A second factor contributing to the varying interpretation of numerical expressions is the type of approximator used in the expression. Precise approximators (e.g., *exactly*) impose a precise interpretation, whereas imprecise approximators (e.g., *roughly, approximately, about*) do the opposite, see (3a). However, due to the tendency of non-round numbers receiving a precise interpretation, it has been pointed out in Sauerland and Stateva (2011)[1] that it is odd to use them together with imprecise approximators, as can be seen in the contrast in (3b).

(3)   a. John made **exactly/roughly 50 cupcakes**.
       b. John made **exactly/?roughly 48 cupcakes**.

While the first and the second factors have received extensive treatment in the literature (a.o., Lakoff 1973; Rips et al. 2007; Krifka 2007, 2009; Sauerland and Stateva 2011; Kennedy 2013; Solt 2014), there is a third factor affecting the

---

[1]What should also be noted with respect to approximators is Geurts' (2006) sharp observation that precise approximators can only modify expressions that already have an exact meaning—while *exactly five sneezes* or *precisely half the cake* are perfectly acceptable expressions, *exactly tall* or *exactly some cookies* are not.

interpretation of numerical expressions which to our knowledge has largely been unexplored, namely, the unit of measurement. Consider (4): the combination of an imprecise approximator and a non-round number is not odd, which stands in contrast to "roughly 48 cupcakes" in (3b). The difference between the targeted expressions is that the unit "cupcake" in (3b) is discrete and the one in (4) "meter" is continuous.

(4)   The tower is **exactly/roughly 48 meters** high.

The current paper examines these three factors in detail, as well as their ways of interaction. The paper is structured as follows. In Sect. 2, we provide a review of related works from theoretical linguistics. In Sect. 3, we report on a corpus-linguistic study with the following main findings: imprecise approximators occur more frequently with round numbers (e.g., *roughly* 50) than with non-round numbers (e.g., *roughly* 48). Furthermore, discrete units occur significantly less frequently than continuous units in the latter combination (e.g., *roughly 48 people* vs. *roughly 48 meters*), which indicates the imprecise nature of the continuous unit. In Sect. 4, we report a rating study testing the naturalness of imprecise approximators in combination with different kinds of numbers and different kinds of units. Our results show both effects by Number and Unit but no interaction between them. Section 5 provides a general discussion and concludes the paper.

Generally speaking, this chapter provides insights into the representation and application of numerical concepts. We focus our research on the usage and interpretation of these concepts in natural language texts, using the results of both a corpus study and a rating study. In our literature review, we summarize different formal models for representing the meaning of numerical expressions, which can be seen as (partial) representations of numerical concepts. In our two studies, we then seek to confirm the qualitative predictions made by these models about the practical usage of such numerical expressions. Our work can be related to the contribution by Gust and Umbach (Chap. 4) who also consider the granularity of interpretation for natural language phrases. While their work targets similarity expressions of varying kinds, we put our focus on expressions that involve concrete numbers. Our experimental rating study can be related to the procedure by Scerrati et al. (Chap. 6) who record binary responses on individual words, while we make use of Likert scale ratings on complete sentences. Finally, the focus on the interpretation of natural language phrases is also investigated by Vernillo (Chap. 8), who uses a theoretical analysis of individual verbs based on image schemata, while we perform a corpus study and a rating study on more complex phrases.

2   Theoretical Background

In this section, we provide a detailed discussion of the three linguistic factors influencing the overall interpretation of numerical expressions, based on the literature. As our concern is on their semantics and pragmatics, we assume a simplified "NumP"

(i.e., number phrase) structure for them consisting of a NumP-modifier (e.g., *exactly*), a Num head (e.g., *fifty*), and an NP complement (e.g., *people*), but are open to alternative syntactic structures.

2.1   Number: Round Versus Non-round

The discussion of round in contrast to non-round numbers is heavily intertwined with the topic of the granularity of scales in which we think. Thinking on a coarse-grained level can be seen as thinking in gross bins. A fine-grained, possibly continuous (i.e., maximally fine-grained) scale is simplified by turning it into a discrete scale with fewer values, therefore coarse-grained thinking means simplified thinking. While these few values are salient and meaningful in the way that we can quickly process and interpret them in a given context, using coarse-grained scales potentially results in less precise reports in measuring contexts.

If we look at scales of different granularity levels such as (5), we will find that round numbers appear both on fine-grained and on coarse-grained scales. This is not the case for non-round numbers—the more coarse-grained a scale becomes, the fewer non-round numbers it contains.

(5)   Scales progressing in steps of 10, 5 and 1 respectively
       a. 0…….....…………….10………………...………………20
       b. 0…….…….5…….…….10….…………….15…….…………20
       c. 0..1..2..3..4..5..6..7..8..9..10..11..12..13..14..15..16..17..18..19..20

Only values on a coarse-grained scale however can represent a whole range of other values; thus, since the values appearing on coarse-grained scales usually are round numbers, round numbers logically allow for an imprecise interpretation. In contrast, non-round numbers do not appear on coarse-grained scales and therefore only lend themselves to a precise interpretation. Thus, one would rather interpret expressions imprecisely that make available an imprecise interpretation than expressions that do not allow such an interpretation. This is why we tend to interpret round numbers imprecisely and non-round numbers precisely.

But what does 'round' really mean? The concept of roundness depends on the context. Solt (2014) speaks of a gradient nature of roundness, meaning that there is a 'more' and a 'less' to roundness: the hierarchical ordering of scales with respect to granularity yields this gradient. For example, 5 can be considered round since it also appears on the more coarse-grained scale (5b), but less round than 10, which appears on an even more coarse-grained scale (5a). In some cases, a number might be considered round if it only has—or is rounded to—two decimal places. In other cases, non-round numbers can take on the same function as round numbers, such as 12 or 24 h in a coarse-grained time scale (see more examples in Krifka 2007). In other words, the availability of an imprecise interpretation of a number does not necessarily depend on it being round; it rather depends on its coarse-grainedness within a system of representation. As our numerical reasoning most commonly makes

use of the decimal system however, which is a base-ten numeral system, round numbers like 10, 100, etc. and simple fractions of them most frequently coincide with coarse-grainedness and are thus more likely to be interpreted imprecisely.

Krifka (2007, 2009) assumes two general pragmatic principles from which he derives (and which shall explain) the RNRI ("Round Numbers, Round Interpretations") phenomenon: (I) weak preference for simple expressions, (II) strict preference for truthful interpretations. The first principle explains why round numbers are used more imprecisely than non-round numbers. The second principle explains why round numbers are interpreted more imprecisely than precisely.

In more detail, Krifka assumes a conditional preference for simple expressions, which explains the approximate usage of round numbers in contexts that do not require high precision. If a speaker has the choice between uttering forty-eight or fifty, he will most likely choose the simpler expression, for reasons of communication efficiency. The preference is conditional in the sense that it can only come into effect if the difference between the two numbers is not relevant in the context (e.g., with specific QUDs or decision problems). Under a precise interpretation, however, the preference cannot come into effect; the speaker does not have the choice between one expression or the other. Krifka models the virtual equivalence between two measure expressions in low-precision contexts in the following way: Under an approximate interpretation, numbers represent ranges which can be characterized by a mean, i.e., the number which the interval is centered around, and a standard deviation, defining the borders of the interval, which also indicates the level of imprecision.[2] Naturally, ranges of two numbers can overlap if the values are close to each other. Two numbers are said to be indistinguishable from each other under an approximate interpretation if the ranges they represent overlap in such a way that their means are within their standard deviations. Under an approximate interpretation, forty-eight could for instance represent the range [46, 47, 48, 49, 50] (having the mean 48 and the standard deviation 2), whereas fifty would represent [48, 49, 50, 51, 52] in that case. Their means are within their standard deviations, so they are considered indistinguishable under this approximate interpretation. However, fifty has the advantage over forty-eight in that it has a simpler form (and is also otherwise more cognitively salient). The speaker thus chooses to utter fifty instead of forty-eight in a context where approximate interpretations are licensed. This also explains why non-round numbers are not interpreted in an approximate way: Once there are several indistinguishable alternatives one could make use of when reporting a measurement, the alternative with the simplest form is chosen, which excludes non-round numbers from the race.

Under a precise interpretation, numbers denote only themselves: forty-eight denotes 48 and fifty 50. The possibility of choosing between alternatives does not arise because their denotations are clearly different.

(6)   a. John made **50 cupcakes**.
       b. John made **48 cupcakes**.

---

[2]More specifically, Krifka models an imprecise number as a normal distribution which is centered around the number. To simplify things, he confines his discussion to a representation in terms of intervals.

Assuming a context which licenses an approximate interpretation, Krifka's model explains the acceptability of (6a) since fifty represents the range [48, 49, 50, 51, 52] which includes 48 and 51. If the context requires a precise interpretation, fifty represents only 50; the usage of this numeral thus would make (6a) false in situations where John made 48 or 51 cupcakes. Similarly, forty-eight in (6b) could represent the range [46, 47, 48, 49, 50] under an approximate interpretation. However, the speaker would have uttered fifty in such a situation, since under an approximate interpretation fifty is indistinguishable from forty-eight, and it is simpler. Thus, forty-eight cannot be interpreted imprecisely here—instead, it must denote solely its own value.

The second principle ought to explain an assumption specific to Krifka's theory. By way of principle (II), the preference for truthful interpretations, Krifka explains why an approximate interpretation of an encountered round number is more sensible than a precise one. Krifka holds the assumption that we prefer an imprecise interpretation of round numbers and therefore usually interpret round numbers imprecisely (an assumption challenged by Ferson et al. 2015). He argues that an imprecise interpretation maximizes the probability of truth of the statement: It is more likely that the value of a reported measurement is in the range of the interval around the reported number (which amounts to an approximate interpretation) than it is likely that the value is the number itself (which amounts to a precise interpretation). And since Krifka also assumes that we follow principle (II), he concludes that the approximate interpretation is the preferred one. On the other hand, an addressee can conclude from an utterance containing the more complex expression that a precise interpretation must have been intended since this is the only context where complex expressions are used—whenever possible, i.e., under an approximate interpretation, the simpler expression (which coincides with round numbers in this case) is chosen over the more complex alternative.

So far, Krifka's argumentation had little to do with a theory of granularity. One might ask however why it is generally the case that round numbers are simpler than non-round numbers. It turns out that the superficial simplicity argument can be reformulated in terms of the scale granularity framework. Krifka points out that it is not just the simplicity of the form of some expression that contributes to whether it is interpreted precisely or imprecisely. Instead, what matters even more is the expression's simplicity in terms of representation. This is where scale granularity becomes important. The simplicity of representation is marked by whether a value is cognitively salient on the scale of reference.

A numerical representation might be perceived as simple (more easily graspable) if it appears on coarse-grained scales of the unit. It becomes clear that the term *simple* here refers to how easily we can process the conveyed bit of information, as in the aforementioned example of time scales {0, 12, 24, 36, 48, …}. Notice that twenty-four is neither simpler than twenty-three in terms of form nor round. It is because of the expression's simplicity of representation and persistence throughout scales of different granularity levels that a speaker might choose twenty-four over twenty-three under an approximate interpretation.

We can conclude that a simple representation promotes an imprecise interpretation because it allows one to reason on a coarse-grained level of scales. Krifka additionally

argues that in many cases, simplicity of expression and simplicity of representation coincide—not coincidentally, but because the frequency of use dictates such a development. Simplification of expressions is a result of an increase of frequency due to their additional approximate use: "salient representations tend to be shorter, and tend to be shortened in language change" (Krifka 2007).

Generally speaking, a characteristic of a round number is that it is simple: self-contained (no infinite decimal places) and conceptually graspable and decodable; it is a number that exists in a simple system of representation (for instance a system of multiples of tens)—the system depends on the context of use. In this paper, we will restrict our empirical analyses to a limited set of (conventionalized) round numbers (e.g., 10-roundness and 5-roundness, which do not need contextual support) in contrast to their non-round close numbers.

2.2 Approximator: Approximate Versus Exact

While we have discussed that (im)precise interpretations of numerals can arise from implicit assumptions about the numbers themselves, there is also an overt means for marking the intended level of precision. Approximators like *exactly, precisely, around,* and *approximately* are classified as hedges (Lakoff 1973): Expressions which modify the certainty, force, or precision implied by statements. Also belonging to this class are expressions like *maybe* or *I assume* (called *shields*), which can modify whole sentences. Approximators are a means of explicitly marking the degree of precision with which a measure expression is to be interpreted, but on a different level, the use of approximators also reveals something about the certainty with which a speaker utters something. The latter is evident if we consider uses of the approximators as speech-act adverbs, e.g., *Roughly speaking, I have 50 students in my class.* We leave it for future studies what differences such sentences have compared to *I have roughly 50 students in my class.*

When a speaker intends to indicate a high certainty about the accurateness of the uttered numeral, they likely use precise approximators. When doing so, the speaker simultaneously decreases the risk of conveying false information, which is higher with an unmodified alternative. In other words, using approximators increases the probability of the truthfulness. Thus, using imprecise approximators can also signal the speaker's uncertainty in addition to imprecision in measuring, which is emphasized in Ferson et al.'s (2015) work.

While Krifka's (2007) work is not concerned with the effect of approximators on numerical expressions, Solt (2014) extends the granularity-based framework to provide an account of these modifying expressions. She also introduces a new formalism for determining truth or falsity of sentences with numerical expressions that includes a contextually determined granularity level. In her analysis, the overt use of approximators in combination with numerals is modeled as a mapping from point-denoting expressions (the bare numerals) to intervals around these expressions.

Explicitly modified numerals thus denote a scalar segment. Solt formally defines the semantics of approximators as in (7):

(7) $[[\text{APPROXIMATOR n}]]^g = (n - \text{gran}'/2, n + \text{gran}'/2)$

For imprecise approximators, gran' is the coarsest possible unit for a granularity level one could choose given the context. For precise approximators, gran' is the finest possible choice of a granularity level given the context. Thus, $[[\text{about 50}]]^{g[\text{gran}'=10]}$ would denote the interval (45, 55) in the appropriate context. It becomes clear that the denotation of a modified measure expression differs from the original numeral in that it (roughly) denotes the range of values halfway between the neighboring values on the coarse-grained level. In formal semantic terms, this complex expression however still is of type "degree" despite not denoting a point.

(8) $[[\text{exactly fifty}]]^{g[\text{gran}'=0.01]} = (50 - 0.01/2, 50 + 0.01/2)$

Notably, Solt's analysis of approximators, as shown in (8), yields as a result that precise approximators can make an expression more imprecise after being combined with the approximator. Although the granularity level is very fine-grained (with gran' being 0.01), the resulting complex expression denotes a more coarse-grained degree than the bare, unmodified numeral, namely, (49.995, 50.005) instead of 50. On the one hand, the analysis of the complex expression is not counterintuitive since in some contexts the usage of a precise approximator does not signal maximal but only increased precision. However, what seems unintuitive is that the bare numeral in contrast can never denote anything more imprecise than the maximally precise point it always denotes. The denotation of the numeral modified by a precise approximator is more imprecise than the denotation of the unmodified numeral. This conflicts with the empirical findings of Ferson et al.'s (2015) study that precise approximators (*exactly* and *precisely*) rather reduce a previously assumed range of imprecision associated with a numeral instead of making numerals more imprecise.

Since Solt's theory does not assume numerals to denote ranges in the first place, there is no way she can model how an approximator can reduce the interval of imprecision that might be associated with a numeral. Thus, this analysis cannot explicitly model situations in which the context favors a default imprecise reading of a numeral while the approximator is used to override this reading. This is only possible within theories that overtly model the imprecision of a numeral such as Krifka who lets unmodified numerals denote ranges under an imprecise interpretation. These representational issues Solt's theory faces due to the assumption of a monosemous exact denotation of numerals might not pose problems in terms of truth-conditional analyses. However, they show that Solt's model is also not entirely optimal as it seems odd to assume that *exactly fifty* denotes a coarse-grained degree while *fifty* does not.

An alternative relates to Lasersohn's theory (1999) of pragmatic halos in which he also proposes an analysis of approximators. Lasersohn takes precise approximators to be narrowing the so-called "pragmatic halos" of an expression: "Suppose, for illustration, that there are two points in time close enough to *i* that the difference

between them and *i* is ignored in context, so that the halo of *three o'clock* is the set {*i*, *j*, *k*}, ordered according the relation of closeness to *i* …. The real effect of *exactly* is on pragmatic halos: we want the pragmatic halo of *exactly three o'clock* to include those elements of the halo of *three o'clock* which are closest to *i* (that is, to the actual time of *three o'clock*), eliminating outlying elements." (Lasersohn 1999: p. 528). In this analysis, precise approximators have no effect on the semantic level, however, they reduce the pragmatic slack with which one may speak and thus have an effect on whether an utterance can be used felicitously or not. This is not the case for imprecise approximators: They are analyzed to have the effect of expanding the denotation of the expression (they are combined with) into its halo. Thus, they have a clear truth-conditional effect in that the resulting denotation is 'enriched' by similar denotations, constituting a set.

Combining Sects. 2.1 and 2.2, a natural question arises as to how numbers interact with approximators. We will not be able to work out a formal analysis here, but focus on the distributional constraints due to the different levels of precision encoded in them.

2.3   Unit: Discrete Versus Continuous

Seeing numbers as part of a mathematical system, we find that at the most basic level, number systems permit the description of quantities by means of expressions consisting of a numeral and a unit, where the unit specifies the scale of measurement. Units can, for instance, be '*people*', '*buildings*', '*chairs*' for discrete quantities, but also '*days*', '*acres*', '*metres*' for continuous quantities.

Accordingly, a numeral can be an integer or real-valued; it furthermore can be expressed in words or numerical digits. Since units measure either discrete or continuous quantities, they can influence the numerals they appear with. Those units measuring discrete quantities restrict the numeral they combine with to the domain of integers. When measuring quantities physically, the numerical expressions used for description are almost always used imprecisely, especially in the case of measuring continuous quantities. Ferson et al. (2015) thus suggest a distinction between the mathematical and the 'real world' interpretation of a numerical quantity. Following this distinction means assuming that in non-mathematical contexts an unmodified scalar number already elicits an interpretation with an interval of imprecision; the expression might refer to any value within this interval. In contrast to this suggestion, however, Ferson et al.'s (2015) empirical study found that participants (who were asked to specify an interval the numbers can stand for) interpreted bare, unmodified numbers precisely in 94% of the time, despite the fact that the expressions were embedded in a natural language context.

What are the effects of units on the distribution and interpretation of number words and expressions? We will provide partial answers to this understudied question in the rest of the paper.

2.4 Summary

In summary, the use and understanding of numerical expressions are subject to influences from both broad discourse contexts and narrow linguistic contexts. In the paper, we will not provide formal analyses for numerical expressions; instead, we focus on the empirical testing of the observations from the literature and the current work. In the following, we will discuss numerical expressions with two goals: First, we will provide empirical (i.e., corpus- and psycholinguistic) evidence for the generalizations related to the distinction between round and non-round numbers. Second, we will provide empirical evidence for the effect of unit in the interpretations of numerical expressions.

3 Corpus Study

3.1 Hypotheses

The aim of the corpus study is, first of all, to support the initial observation made, namely that round numbers seem to appear more frequently in natural language contexts than expected if they only had a precise usage. If confirmed, this more frequent appearance is taken as support for the claim that round numbers, in addition to denoting their own values, are used imprecisely due to context (e.g., when imprecision prevails over precision, or when the speaker is uncertain about the actual precise values). Their additional use for this purpose would explain the prevalence of round numbers throughout natural language data. Furthermore, the analysis has been conducted to shed light on the distribution of approximator (null/precise/imprecise), numeral (round/non-round) and unit (discrete/continuous), as well as possible patterns in their conjoint appearance.

Based on the theoretical considerations in Sect. 2, we started with the following hypotheses where π_1 denotes the probability of the number i occurring in natural language communication:

(9)   H0: $\pi_1 = \pi_2 = \ldots = \pi_{500}$
         H1: $\pi_1 \neq \pi_2 \neq \ldots \neq \pi_{500}$

In the null hypothesis H0, each numeral is assumed to appear with an equal probability in the corpus. The corpus study restricts numerical analysis to numerals in the range between 1 and 500, hence the notation above. Say the probability of appearance of each numeral is 1/500, then we expect round numbers (i.e., numbers ending with a 0 or 5) to appear 20 percent of the time (100 out of the 500 numbers are round) whereas non-round numbers should appear 80 percent of the time (the remaining 400 out of 500 numbers are non-round).

Our first hypothesis is captured in the H1: We expect that the probability of appearance is not equal for every numeral. More specifically, related to H1, we assume that round numbers appear more often than expected (i.e., >20%).

Secondly, we assume that the default interpretation of numerals in general is precise, following Ferson et al. (2015) and the findings in their study (and contrary to Krifka 2007). As a consequence, a precise interpretation often does not have to be signaled explicitly whereas imprecise approximators are needed to signal an intended imprecise interpretation. Thus, our second hypothesis is that precise approximators appear less frequently than imprecise approximators.

Thirdly, in terms of combinations of approximators and numerals, let us recall example (3b) or (10a) from Sauerland and Stateva (2011), which they take to be odd. Since imprecise approximators usually signal a coarse granularity level, the appearance with a non-round number (which only appears on more fine-grained scales) strikes the reader as peculiar. We will therefore expect that imprecise approximators tend to appear with round numbers.

(10) a. # What John cooked were approximately 49 tapas.
      b. The rope is approximately 49 metres long.

Furthermore, theoretical accounts so far mainly focused on the interaction between approximators and numerals. Ferson et al. (2015) examined a potential influence of the unit on the interpreted imprecision of a numeral, a hypothesis that was not supported by the results of their study. To our knowledge, little attention has been paid to the potential interaction between unit, approximator, and numeral, see for instance, (10b). Whereas (10a) is odd to the reader, this oddity disappears in (10b), which is completely natural. This can be attributed to the fact that the continuous unit implies that 49 m can already be used imprecisely (49 is round compared to 48.7) whereas this is not the case for discrete numbers (49 is the most precise possible in this case and has no imprecise reading). The results of the corpus study will also be inspected with respect to this effect.

3.2   Methods

The study was based on the Leipzig Wortschatz corpus (Goldhahn et al. 2012), containing 1 million English sentences sourced from online news reports and general web crawling results. The corpus was searched for numerical expressions in the Approximator-Number-Unit fashion. The code was written in python and is publicly available online (https://github.com/lbechberger/CorpusStudyNumerals). The matches were analyzed with respect to the following variables:

(11) Variables
      a. Approximator: precise, imprecise, null
      b. Number: round, non-round
      c. Unit: discrete, continuous

Counts kept track of the different combinations. Numbers were counted as round if they ended with a zero or five; we only used integer numbers (excluding decimal numbers) in the analysis. We only included number words up to five hundred in the counts. The categories for the approximator matched for the following words:

(12)   Categories of approximators (Approx.)
       a. Precise Approx.: ['exactly', 'precisely']
       b. Imprecise Approx.: ['about', 'approximately', 'roughly', 'around', 'round about', 'roughly around', 'some'[3]]
       c. Asymmetrical Approx.: ['more than', 'nearly', 'over', 'almost', 'approaching', 'below', 'above', 'fewer than', 'less than', 'at most', 'at least', 'close to', 'near to', 'up to', 'as high as', 'as low as', 'not quite']
       d. Null Approx.: every expression preceding a numeral that does not match the words above

Asymmetrical approximators (based on Ferson et al.'s (2015) list of approximators used in his study) were not included in the statistical analysis. Yet, they were also matched to obtain an estimate of the frequency of their usage and have a more accurate account of the unmodified versus modified numerals ratio. Their appearance with either round or non-round numbers was neither recorded nor analyzed (although asymmetrical approximators, a.k.a. comparatives, are also a subject of debate in current accounts of imprecision (Solt 2014)). The unit was first matched as any word following the numeral and subsequently evaluated using WordNet (Princeton University 2010) for whether it belonged to one of the following categories:

(13)   Categories of units
       a. continuous: ['time period', 'time unit', 'linear unit', 'magnitude relation', 'monetary unit', 'unit of measurement']
       b. discrete: ['organism', 'human activity', 'group', 'location', 'transport', 'material']

All numerals occurring with matches that did not belong to any of the categories have been excluded; the remaining matches were used for the analysis. The data consequently had the nature of frequency counts of the aforementioned *(Approximator)-Number-Unit* sequences and of the respective counts of approximators, numerals, and units separately. The analysis consisted of testing the match counts against their expected frequency: The main hypothesis, that the frequency of round numbers is different from their expected frequency, was tested for significance using the Binomial Test. The effects in the Number (roundness) * Approximator and Unit * Approximator contingency tables were tested using the χ^2 Test.

---

[3]Such as *Some 50 students joined the protest*.
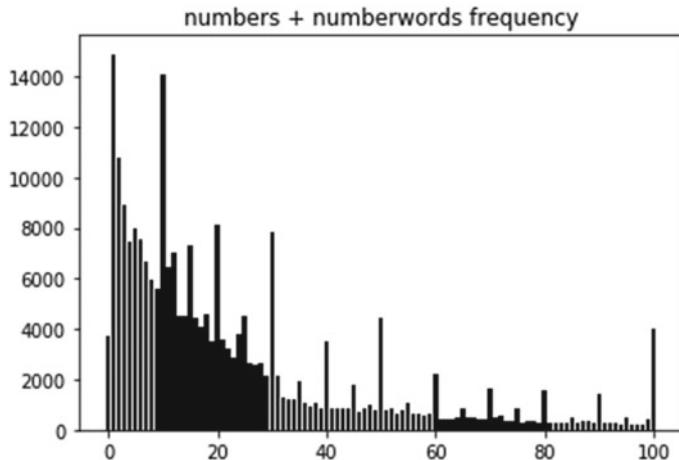
Fig. 1 Numeral counts from 0 to 100
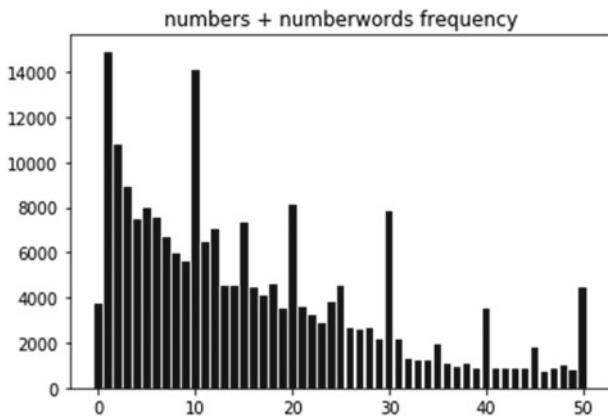
Fig. 2 Numeral counts from 0 to 50

3.3 Results and Interpretation

As can be seen from Figs. 1 and 2, there are "spikes" of counts for round numbers (also visible in the range between 0 and 100) already indicating a marked appearance of round numerals in the corpus. The general distribution (a few numbers with very high counts and a tail to the right) suggests that numeral occurrences seem to follow a power law distribution, specifically one related to Benford's law (Benford 1939). The extraordinarily high count for the numeral 1 can be explained by the frequent usage of the number word in many contexts (e.g., 'He had one goal.', 'A government has the energy for only so many fights at one time.', etc.).

Table 1 Frequency counts of matches in the corpus: Approximator * Number (roundness) * Unit

n = 182,895	Discrete		Continuous		
	Round	Non-round	Round	Non-round	
Precise	4	7	15	36	(62)
Imprecise	2,975	**423**	3,217	**2,081**	(8,696)
Null	21,215	64,990	30,535	57,397	(174,137)
	(24,194)	(65,420)	(33,767)	(59,514)	

Table 2 Frequency counts of matches in the corpus: Number (roundness) * Approximator

n = 182,895	Precise	Imprecise	Null	
Round	19	**6,192**	51,750	(57,961)
Non-round	43	**2,504**	122,387	(124,934)
	(62)	(8,696)	(174,137)	

182,895 of the matched numerals were used for the analysis (another 369,384 in that range were discarded due to unit constraints). The null hypothesis thus expects 36,579 of these numerals to be round and 146,316 numerals to be non-round.

Generally, as in Table 1, we observe the following tendencies: First, non-round numbers appear, in absolute terms, more often than round numbers. Second, unmodified numerals appear most frequently with a count of 174,137, followed by numerals modified by an imprecise approximator (8,696 counts) and lastly, numerals modified by precise approximators (62 matches). Third, numerals with discrete units (89,614 counts) appear almost as often as numerals with continuous units (93,281 counts), with a ratio of approximately 0.49/0.51.

More specifically, our findings are stated as follows, see Table 2. First, round numbers appear more frequently than expected. As we can read from the tables, a total of 57,961 (as opposed to the expected 36,579) round numbers and a total of 124,934 (as opposed to the expected 146,316) non-round numbers were counted. Instead of an expected 0.2/0.8 ratio, we found a ratio of approximately 0.32/0.68. This effect is particularly pronounced if the numerals appear with a continuous unit—the ratio between round and non-round numbers is roughly 0.36/0.64 there. Binomial testing reveals that this is a significant departure from the expected frequency ($p < 0.01$, one-sided).

Second, imprecise approximators appear more frequently than precise approximators. Table 3 shows a total count of 8,696 imprecisely modified numerals as opposed to the few 62 occurrences of precisely modified numerals in the given range. This undoubtedly supports our assumption that the default interpretation of numerals is precise which makes imprecise approximators an important tool to signal that the imprecise interpretation is intended, whereas precise approximators are unnecessary most of the time.

Table 3 Frequency counts of matches in the corpus: Unit * Approximator

n = 182,895	Precise	Imprecise	Null	
Discrete	**11**	3,398	86,205	(89,614)
Continuous	**51**	5,298	87,932	(93,281)
	(62)	(8,696)	(174,137)	

Table 4 Breakdown of Table 1 with respect to Unit

	Discrete		Continuous		
Imprecise	Round	Non-round	Round	Non-round	
	2975	**423**	3217	**2081**	(8696)

Third, imprecise approximators tend to appear with round numbers, especially if the unit is discrete. This is one of the most impressive results from the study: Even though generally and in absolute terms, non-round numbers occur more often than round numbers, we can read from Table 2 that if numerals occur with an imprecise approximator, the proportions are almost swapped. Even in absolute terms, imprecisely modified round numerals occur more often than imprecisely modified non-round numerals. This represents strong evidence for our hypothesis that imprecise approximators predominantly appear with round numbers. The deviations from the expected frequencies in Table 2 were significant using the χ^2 Test, i.e., χ^2 (df = 2, n = 18,2895) = 6585.259, $p < 0.01$, $\varphi_c = 0.19$. Conversely, this finding can be framed in terms of the infrequent appearance of imprecise approximators with non-round numbers (see Sauerland and Stateva's (2011) oddity example (10a) mentioned). Arguably, 2,504 occurrences of non-round numerals appearing with imprecise approximators is still a substantial count. Resolution however comes from looking at Table 4 where a further breakdown of the data with respect to the unit category is presented:

We see that this effect is particularly strong if we are looking at the discrete domain: There were 2,975 occurrences of the imprecise approximator-round numeral combination, whereas only 423 non-round numerals appeared with an imprecise approximator there (roughly an impressing 0.88/0.12 ratio). This is in line with Sauerland and Stateva's (2011) observation about imprecise approximators occurring with non-round numbers. In contrast, in the continuous domain, this effect vanishes for the most part (compare (10b)). This is also reflected in our counts: Although it is still the case that imprecise approximators occur more often with round numbers in this condition, the count for imprecise approximators appearing with non-round numbers is almost equally high and in absolute terms not negligible. This indicates that the oddity of imprecise approximators appearing with non-round numbers is drastically reduced if these numbers are continuous. We have thus encountered evidence for the claim that the unit has an effect on the co-occurrence behavior of approximators and numerals.

Last but not least, precise approximators tend to appear with continuous units. We see that if precise approximators appear at all, they tend to co-occur with continuous units (51 occurrences with continuous quantities vs. 11 occurrences with discrete quantities, see boldfaced numbers in Table 3). This makes sense to the extent that for continuous quantities, the precise interpretation is not trivial. These observations, however, should be taken with a grain of salt as we did not have many occurrences of precise approximators overall.

(14)   a. Trump announced his candidacy for the Republican nomination **exactly three months** ago.
       b. Belgium's federal prosecutor's office says authorities have so far made **(?exactly) three arrests** linked to the deadly attacks in Paris.

While *exactly* adds nothing to the already precise interpretation of (14b), in (14a), it makes a contribution to the interpretation of the numeral. Since the used numeral in (14a) can never be entirely accurately describing the actual time span between Trump's announcement and the report, the degree of accurateness needs to be marked explicitly to indicate "how precisely" the expression is meant. In (14a), one can assume that the speaker intended an interpretation accurate to the day (i.e., the report was made on the same date of the third subsequent month). Unless the numeral is of special interest, *exactly* in (14b) in contrast, appears redundant.

4   Psycholinguistic Experiment

To investigate the effect of the unit on the acceptability of numerical expressions, we tested English numeral expressions using a 2×2 factorial design, with the factors *Number* (round vs. non-round) and *Unit* (discrete vs. continuous).

4.1   Materials and Predictions

We used 24 different matrix sentence items, each in four conditions, see the Appendix for the entire list of the items. The experimental items were constructed under the following objective: The setout was to choose sentences containing the sequence *imprecise approximator—round number*, which has been motivated to evoke no perception of oddity. The sentences were picked from the Leipzig Wortschatz corpus. Before selecting the sentences, we determined the round numbers that they ought to contain. For this, 12 round numbers were randomly chosen in the range from 10 to 1000. This yielded the numbers 10, 60, 70, 100, 350, 400, 700, 750, 800, 900, 950, 1000. We then scanned the corpus for sentences containing imprecise approximators (*about, around, approximately* and *roughly*, six occurrences each) and the randomly chosen round numbers that would appear with either discrete or continuous

units, resulting in equally many sentences for both the 'discrete' and the 'continuous' condition.

Based on the experimental items for the *round* conditions, we created their non-round counterparts by changing the round number of each sentence into a close-by non-round one. This way we ensured that the non-round number would appear in a plausible context and linguistic environment. The oddity could thus only arise from the pairing of a non-round number with an imprecise approximator. The four conditions are exemplified in (15).

(15) a. **r-disc**: As of then, **about 60 Cubans** had arrived in the Yucatan coast in 2015.
      b. **r-cont**: Brigham City is **about 60 miles** north of Salt Lake City.
      c. **nr-disc**: As of then, **about 61 Cubans** had arrived in the Yucatan coast in 2015.
      d. **nr-cont**: Brigham City is **about 61 miles** north of Salt Lake City.

Additionally, we used 48 filler items as distractors, which were news report sentences of comparable length that we also sourced from the Leipzig Wortschatz corpus. We did not revise these, as the pragmatic difference at focus is subtle and thus fillers containing ungrammatical or odd phrases would be inappropriate.

(16) *The drug investigation began **in August 2013** at Edwards Air Force Base in California.*

Based on Sauerland and Stateva's (2011) observation that non-round numbers are odd with imprecise approximators and our corpus-linguistic finding that this effect is stronger with discrete units than continuous units, we had the following predictions: First, there will be a main effect of Number. More specifically, the condition "nr-disc" will be rated worse than "r-disc", and the condition "nr-cont" will be rated worse than "r-cont". These predictions are in accordance with the oddity suggested by Sauerland and Stateva. Due to the observations made in the corpus study, we included a second prediction, namely, there would be a main effect of Unit and possibly an interaction between Number and Unit due to a stronger worsening effect with discrete than with continuous units.

We used a Latin Square design, that is, each participant read one set of 72 sentences in total. As seen above, the participants' attention was directed towards the phrases of interest by marking the relevant phrase visually, both in experimental and in filler items.[4] For the filler items, the marked phrases were mostly DP's or PP's (i.e., determiner phrases or prepositional phrases).

---

[4]We highlighted the critical phrases, as in the pretest without doing so, the subjects did not distinguish the conditions, raising the question whether it showed no evidence for the effect of unit, or whether it was due to methodological issues.

4.2 Procedure and Participants

The experiment was set up with Ibex Farm (spellout.net/ibexfarm/), a website that provides free hosting for online psycholinguistic experiments. Experimental data was gathered using Amazon MTurk, a crowdsourcing platform where human intelligence tasks (HITs) can be carried out by participants who receive compensation for each HIT completed. Requesters were provided the link to the experiment and compensated with $4. Native English-speaking workers on Amazon MTurk (N = 72) signed informed consent and participated in the study.

Before entering the experimental phase, participants first completed a practice session where 12 practice items were to be rated. During the experimental phase, they first read an entire sentence and then were asked to rate the naturalness of the underlined phases (which were shown again separately) on a 7-point Likert scale (1 = unnatural, 7 = natural).

4.3 Data Analysis and Results
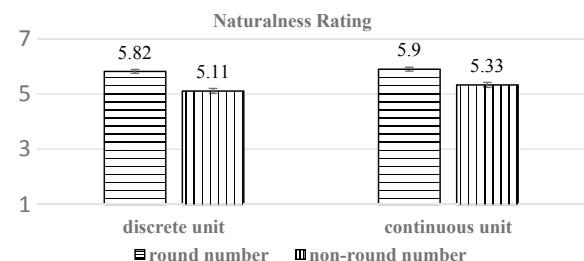
The descriptive statistics is provided in Table 5 and visualized in Fig. 3. As can be seen in the table, descriptively, the "r-cont" condition received the highest mean rating, whereas the "nr-disc" condition received the lowest mean rating. The standard deviation was also highest for the "nr-disc" condition, indicating an overall lower consistency in ratings for this condition.

We analyzed the data using R. All analyses were performed using mixed effects linear regression models; the models were constructed using the *lme4* package in R

Table 5 Mean naturalness ratings (1 = unnatural, 7 = natural) and standard deviations (SDs) and standard errors (SEs)

Conditions	Mean	SD	SE
r-disc	5.82	1.47	0.07
r-cont	5.90	1.39	0.07
nr-disc	5.11	1.77	0.09
nr-cont	5.33	1.70	0.09

Fig. 3 Naturalness ratings of the experiment

(Baayen et al. 2008; Bates et al. 2012). All contrasts of interest were sum coded and included as fixed effects in the model. The reported model is the maximal model that converged. The model included *Number* and *Unit* (with interaction term) as fixed effects. Furthermore, we included random intercepts for subjects, items, and stimuli order, as well as random by-subject slopes for the effects of Number and Unit (and their interaction).

We found a significant main effect of *Number* (t = 4.15, p < 0.0001). Tukey's HSD for multiple comparisons of means indicates that round numbers were rated significantly more natural than non-round numbers with both continuous (t = 3.96, p < 0.005) and discrete (t = 3.84, p < 0.005) units. Furthermore, we found a significant effect of *Unit* (t = 2.11, p < 0.05) in that continuous conditions were rated better than discrete conditions. However, there is no interaction between the two factors, which suggests that the effect of neither factor is influenced by the presence or absence of the other.

In this study, we were able to confirm our first predictions about the effect of Number and Unit. We will leave the reason for the lack of an interaction for future studies.

5   Discussion and Conclusion

In this paper, we tried to gain insight into our understanding and interpretation of numerical expressions with regard to questions such as whether numbers are imprecise at the semantic level.

5.1   Numbers and Number Concepts

We must keep in mind that the development of the number system as we know it now has been a process of cultural construction and added knowledge over generations and centuries of historical time. When analyzing how we interpret numerical expressions in natural language contexts, insight might be provided by looking at the innate numerical concepts humans (and non-human animals) are equipped with for reasoning quantitatively.

Our understanding of number proceeds from concepts that do not conform to the structure and characteristics of the natural numbers (Rips et al. 2007). Two main mechanisms for quantitative reasoning have been identified for numerical ability in infants and non-human animals: On the one hand, a system works with internal analog magnitudes—perhaps some type of continuous strength or activation—which is a linear function of the input. On the other hand, infants' skills for quantitative reasoning may also draw on discrete and distinct representations of objects that are kept in short-term memory—however only less than four items can be represented this way.

Explained shortly, a mental (i.e., internal analog) magnitude is an internal representation of a quantity—this can be the cardinality of a set, but also duration, length, or volume of whatever is registered by the organism. What is special about this representation is that it is assumed to represent an objective magnitude in a direct linear relationship—in that it constitutes a continuous quantity (e.g., activation strength) represented mentally that adjusts to achieve a measure of a quantity. It is thus suggested that mental magnitudes share the formal properties of real numbers (Gallistel and Gelman 2005). However, analog magnitude representations are noisy, and the noise linearly increases the bigger the quantities become. This means, the bigger the measured values, the more imprecise the representation.[5] Analog magnitude representations of large sets are thus only approximate; they are a coarse representation, contrasting with the precision associated with natural numbers.

The other mechanism makes for an infant's ability to predict the total number of objects in small sets (less than 4) and might be considered conceptually closer to the elaborate concept we have of integers.[6] It depends on attentional or short-term memory mechanisms that represent individual objects as distinct entities. For each object, there is a distinct representation within the four-object capacity limit. A set exceeding three items cannot be held in the infants' short-term memory (Carey 2004).

Many psychologists believe that full-fledged mathematical thinking mainly originates from these two innate concepts that are also shown to be existent in non-human animals. Although other researchers argue that these abilities do not seem to be adequate prerequisites for forming the mathematical concept we have of numbers within a number system (see Rips et al. 2007 for a discussion of this issue), they are still shown to have relevance in quantitative and even arithmetical reasoning (Gallistel and Gelman 2005). In specific, analog magnitudes are shown to play a role in arithmetical computations: comparison of two values, and also addition and subtraction (Carey 2004). Indeed, if analog magnitude representations are made use of in mathematical contexts, which would most of all require high precision representations, it is likely that they are also employed when encountering numerical expressions in a natural language context.

How, however, do these mechanisms play into the interpretation of numerical expressions in natural language, if they do so at all? Krifka (2007) argues that the existence of these two distinct systems of representation provides plausibility for both an exact and an approximate interpretation of numerals since they work in parallel and are not hierarchically ordered in any way. Which one of the two is the "original" meaning of a numeral is not settled by this argumentation, it might even be that there is none and that both interpretations are equally prevalent. All the findings in developmental research however do not comprise or imply an inherent distinction

---

[5]In this, mental magnitudes follow Weber's law, according to which the discriminability of two values is a function of their ratio: The bigger the physical magnitudes (and consequently the analog magnitudes), the harder discrimination between pairs of values that are separated by the same absolute difference becomes.

[6]Short-term object representations do have the discreteness of natural numbers; however, they do not form a set representation of the tracked objects and, consequently, cannot represent cardinality. This in turn is represented by mental magnitudes.

of round vs. non-round numbers with respect to impreciseness. Thus, at least the imprecise interpretation of round numbers (not the general imprecise representation of quantities) seems to be a phenomenon "on top of" the basic interpretation of numerals, which likely only started to develop after the formation of more elaborate mathematical systems.

5.2 Contributions and Outlooks of the Current Study

In the current study, we provide a critical discussion of numerical expressions based on the recent formal (compositional) semantic literature, focusing on the imprecise and precise interpretation of numerical expressions. While the interpretation of numerical expressions depends on both broad discourse context and narrow linguistic context, we only dealt with the latter. Our corpus and experimental studies show that the interpretation of numerical expressions is subject to the kind of numbers, the kind of units, as well as whether and what approximators co-occur with them.

It is to note that the results we obtained in our study are certainly contingent on, for example, the specific corpus study or experimental design, the specific numerals (i.e., 0–500) we used, and the specific contexts they occurred (in our case, naturally occurring contexts instead of made-up contexts as in usual experimental works), thus, whether and to what extent they apply to numerical expressions in general need to be investigated in further studies. Furthermore, approximators might differ among themselves. For example, even within the imprecise category, *roughly* and *some* as in *some 50 people* might have syntactic, semantic, or pragmatic differences, which we were not able to handle here. The same holds for Unit which might differ in terms of aspects other than discreteness. Another question for future studies is how the interpretation of numerical expressions is manipulated by broad context (such as QUD, decision problems, developmental, or individual differences, purely information exchanging vs. strategic communication, counting vs. measuring contexts, to just name a few parameters). Despite of this, we believe that the method and the findings of the paper have made further steps to understanding numerical concepts and related concepts that they modify.

Appendix: Test Items of the Experiment
(I./C. For Item/Condition)

I./C.	Sentence
1/1	People from <u>about 10 municipalities</u> around the wrecked plant still cannot go home.
1/2	<u>About 10 hours</u> after the terror attack, someone approached police officers near a synagogue and started shooting.
1/3	People from <u>about 9 municipalities</u> around the wrecked plant still cannot go home.
1/4	<u>About 9 hours</u> after the terror attack, someone approached police officers near a synagogue and started shooting.
2/1	As of then, <u>approximately 60 Cubans</u> had arrived in the Yucatan coast in 2015.
2/2	Brigham City is <u>approximately 60 miles</u> north of Salt Lake City.
2/3	As of then, <u>approximately 61 Cubans</u> had arrived in the Yucatan coast in 2015.
2/4	Brigham City is <u>approximately 61 miles</u> north of Salt Lake City.
3/1	Reportedly, <u>around 70 members</u> of the security forces have been killed in attacks blamed on the PKK.
3/2	BBC Travel says it is currently taking <u>around 70 minutes</u> to get through the traffic.
3/3	Reportedly, <u>around 69 members</u> of the security forces have been killed in attacks blamed on the PKK.
3/4	BBC Travel says it is currently taking <u>around 69 minutes</u> to get through the traffic.
4/1	Experts say present-day cars have <u>about 100 computers</u> on board, and that will double in a few years.
4/2	The helicopter was <u>about 100 feet</u> off the ground when it went out of control and smashed into a parking lot.
4/3	Experts say present-day cars have <u>about 101 computers</u> on board, and that will double in a few years.
4/4	The helicopter was <u>about 101 feet</u> off the ground when it went out of control and smashed into a parking lot.
5/1	The tornado hit on Monday morning, leaving <u>around 350 homes</u> destroyed.
5/2	A hard-charging fire spread across <u>around 350 acres</u> in a matter of hours, the county fire department said.
5/3	The tornado hit on Monday morning, leaving <u>around 349 homes</u> destroyed.
5/4	A hard-charging fire spread across <u>around 349 acres</u> in a matter of hours, the county fire department said.
6/1	In total, <u>about 400 competitors</u> are entered in this year's Dakar Rally, in cars, trucks, motorbikes and quad bikes.
6/2	The Dutch designed Guyana's drainage system <u>about 400 years</u> ago.

6/3   In total, <u>about 401 competitors</u> are entered in this year's Dakar Rally, in cars, trucks, motorbikes and quad bikes.

6/4   The Dutch designed Guyana's drainage system <u>about 401 years</u> ago.

7/1   Nigerian armed forces attacked Boko Haram's headquarters and killed <u>about 700 people</u>, including its leader.

7/2   It was confirmed that there was <u>about 700 tons</u> of the deadly chemical stored at the warehouse that blew up.

7/3   Nigerian armed forces attacked Boko Haram's headquarters and killed <u>about 699 people</u>, including its leader.

7/4   It was confirmed that there was <u>about 699 tons</u> of the deadly chemical stored at the warehouse that blew up.

8/1   The run is taking place in <u>approximately 750</u> communities across Canada.

8/2   The storm was currently situated <u>approximately 750</u> miles east of the Leeward Islands, the National Hurricane Center said.

8/3   The run is taking place in <u>approximately 751</u> communities across Canada.

8/4   The storm was currently situated <u>approximately 751</u> miles east of the Leeward Islands, the National Hurricane Center said.

9/1   Russian military and security forces have an inventory of <u>approximately 800 drones</u>, all believed unarmed.

9/2   South Indian Lake is <u>approximately 800 kilometres</u> north of Winnipeg.

9/3   Russian military and security forces have an inventory of <u>approximately 799 drones</u>, all believed unarmed.

9/4   South Indian Lake is <u>approximately 799 kilometres</u> north of Winnipeg.

10/1   Last year in Manitoba <u>approximately 900 people</u> were diagnosed with colorectal cancer.

10/2   The fire, which is being driven by the wind, is <u>approximately 900 acres</u> in size.

10/3   Last year in Manitoba <u>approximately 901 people</u> were diagnosed with colorectal cancer.

10/4   The fire, which is being driven by the wind, is <u>approximately 901 acres</u> in size.

11/1   The study included an analysis of <u>roughly 950 papers</u> and reports on fracking and water supplies.

11/2   The bus plunged near Joinville, <u>roughly 950 kilometres</u> southwest of Rio de Janeiro.

11/3   The study included an analysis of <u>roughly 949 papers</u> and reports on fracking and water supplies.

11/4   The bus plunged near Joinville, <u>roughly 950 kilometres</u> southwest of Rio de Janeiro.

12/1   <u>Roughly 1,000 residents</u> in the area of a resort town remain under evacuation orders.

12/2   The plane in which he was flying crashed in the Atlantic Ocean, <u>roughly 1,000 miles</u> off course.

12/3   <u>Roughly 1,001 residents</u> in the area of a resort town remain under evacuation orders.

12/4   The plane in which he was flying crashed in the Atlantic Ocean, <u>roughly 1,001 miles</u> off course.

13/1   Authorities then tracked down <u>about 10 men</u> who'd previously been questioned and asked for DNA samples.

13/2   The two car bombs in Baghdad went off <u>about 10 minutes</u> apart late Saturday in the Karrada district.

13/3   Authorities then tracked down <u>about 11 men</u> who'd previously been questioned and asked for DNA samples.

13/4   The two car bombs in Baghdad went off <u>about 11 minutes</u> apart late Saturday in the Karrada district.

14/1   <u>Around 60 firefighters</u> rushed to Adair Tower in North Kensington after a blaze in a two-room apartment.

14/2   According to the defense official, Vietnam has reclaimed <u>around 60 acres</u> of land since 2009.

14/3   <u>Around 59 firefighters</u> rushed to Adair Tower in North Kensington after a blaze in a two-room apartment.

14/4   According to the defense official, Vietnam has reclaimed <u>around 59 acres</u> of land since 2009.

15/1   <u>Approximately 70 people</u> have left Norway to become foreign fighters in Syria or Iraq.

15/2   <u>Approximately 70 miles</u> southeast of the camping resort, a tornado raked Coal City and damaged several subdivisions.

15/3   <u>Approximately 71 people</u> have left Norway to become foreign fighters in Syria or Iraq.

15/4   <u>Approximately 71 miles</u> southeast of the camping resort, a tornado raked Coal City and damaged several subdivisions.

16/1   <u>Around 100 buildings</u> in the city have been flooded, prompting calls for aid from residents.

16/2   Temperatures hovered <u>around 100 degrees</u>, raising concerns of wildfire.

16/3   <u>Around 99 buildings</u> in the city have been flooded, prompting calls for aid from residents.

16/4   Temperatures hovered <u>around 99 degrees</u>, raising concerns of wildfire.

17/1   Officials asked that the <u>roughly 350 people</u> who live in the area known as Morgan's Point remain inside their homes.

17/2   The mine is located <u>roughly 350 kilometres</u> from Mali's main northern city of Gao.

17/3   Officials asked that the <u>roughly 351 people</u> who live in the area known as Morgan's Point remain inside their homes.

17/4   The mine is located <u>roughly 351 kilometres</u> from Mali's main northern city of Gao.

18/1   This year's conference attracted <u>roughly 400 activists</u> from countries including China, Thailand, and Nepal.

18/2   The area spanning <u>roughly 400 yards</u> had been closed since early Wednesday morning.

18/3   This year's conference attracted <u>roughly 399 activists</u> from countries including China, Thailand, and Nepal.

18/4   The area spanning <u>roughly 399 yards</u> had been closed since early Wednesday morning.

19/1   The mayor said <u>about 700 families</u> have voluntarily left the area and more were applying for relocation help.

19/2   The last recorded radar return showed the plane at an altitude of <u>about 700 feet</u> not far away from the crash site.

19/3   The mayor said <u>about 701 families</u> have voluntarily left the area and more were applying for relocation help.

19/4   The last recorded radar return showed the plane at an altitude of <u>about 701 feet</u> not far away from the crash site.

20/1   <u>Roughly 750 people</u> drowned when their trawler sank between Libya and southern Italy.

20/2   The redesigned Ford F-150 pickup is <u>roughly 750 pounds</u> lighter than the older truck.

20/3   <u>Roughly 749 people</u> drowned when their trawler sank between Libya and southern Italy.

20/4   The redesigned Ford F-150 pickup is <u>roughly 749 pounds</u> lighter than the older truck.

21/1   The prison complex holds <u>approximately 800 inmates</u> in separate male and female lockups.

21/2   The Ryan Aeronautical struck a tree and crashed on a golf course <u>approximately 800 feet</u> from the runway.

21/3   The prison complex holds <u>approximately 801 inmates</u> in separate male and female lockups.

21/4   The Ryan Aeronautical struck a tree and crashed on a golf course <u>approximately 801 feet</u> from the runway.

22/1   The United States alone has flown <u>roughly 900 combat missions</u> over Iraq since August.

22/2   The National Hurricane Center said that Storm Ida was centered <u>roughly 900 miles</u> west of the Cape Verde Islands.

22/3   The United States alone has flown <u>roughly 899 combat missions</u> over Iraq since August.

22/4   The National Hurricane Center said that Storm Ida was centered <u>roughly 899 miles</u> west of the Cape Verde Islands.

23/1   An extended Bundeswehr mission in the Mediterranean could in future involve <u>around 950 soldiers</u>.

23/2   Hpakant, <u>around 950 kilometers</u> northeast of Myanmar's biggest city, Yangon, is the industry's epicenter.

23/3   An extended Bundeswehr mission in the Mediterranean could in future involve <u>around 951 soldiers</u>.

23/4   Hpakant, <u>around 951 kilometers</u> northeast of Myanmar's biggest city, Yangon, is the industry's epicenter.

24/1   The warning was for <u>around 1,000 customers</u> in the Hamlin area.

24/2   A handful of prior oil spills have taken place on the line, the largest <u>around 1,000 gallons</u>.

24/3   The warning was for <u>around 999 customers</u> in the Hamlin area.

24/4   A handful of prior oil spills have taken place on the line, the largest <u>around 999 gallons</u>.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

Bates, D. M., Maechler, M., & Bolker, B. (2012). *lme4*: Linear mixed-effects models using S4 classes (R Package Version 0.999999–0). https://cran.r-project.org/web/packages/lme4/index.html.

Benford, F. (1939). The law of anomalous numbers. *Proceedings of the American Philosophical Society, 78,* 551–572.

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus, 133*(1), 59–68.

Ferson, S., O'Rawe, J., Antonenko, A., Siegrist, J., Mickley, J., Luhmann, C. C., et al. (2015). Natural language of uncertainty: Numeric hedge words. *International Journal of Approximate Reasoning, 57*(1), 19–39.

Gallistel, C. R., & Gelman, R. (2005). Mathematical cognition. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambrigde handbook of thinking and reasoning* (pp. 559–588). Cambridge University Press.

Geurts, B. (2006). Take five. In S. Vogeleer & L. Tasmowski (Eds.), *Non-definiteness and plurality* (pp. 311–329). Amsterdam: Benjamins.

Goldhahn, D., Eckart, T., & Quasthoff, U. (2012). Building large monolingual dictionaries at the Leipzig Corpora collection: From 100 to 200 languages. In *Proceedings of the 8th International Language Resources and Evaluation* (LREC'12).

Kennedy, C. (2013). A scalar semantics for scalar readings of number words. In I. Caponigro & C. Cecchetto (Eds.), *From grammar to meaning: The spontaneous logicality of language* (pp. 172–200). Cambridge University Press.

Krifka, M. (2002). Be brief and vague! And how bidirectional optimality theory allows for verbosity and precision. In D. Restle & D. Zaefferer (Eds.), *Sounds and systems. Studies in structure and change. A festschrift for theo vennemann* (pp. 439–458). Mouton de Gruyter, Berlin.

Krifka, M. (2007). Approximate Interpretations of NumberWords: A Case for Strategic Communication. In G. Bouma, I. Krämer & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 111–126). Royal Netherlands Academy of Arts and Sciences.

Krifka, M. (2009). Approximate interpretations of number words: A case for strategic communication. In E. Hinrichs & J. Nerbonne (Eds.), *Theory and evidence in semantics* (pp. 109–132). Stanford: CSLI Publications.

Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic, 2*(4), 458–508.

Lasersohn, P. (1999). Pragmatic halos. *Language, 75*(3), 522–551.

Musolino, J. (2004). The semantics and acquisition of number words: Integrating linguistic and developmental perspectives. *Cognition, 93,* 1–41.

Princeton University. (2010). *About WordNet*.

Rips, L. J., Bloomfield, A., & Asmuth, J. (2007). From numerical concepts to concepts of number. *The Behavioral and Brain Sciences, 31*(01), 623–642.

Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. OSU Working Papers in Linguistics, 49, 91–136.

Sauerland, U., & Stateva, P. (2011). Two types of vagueness. In P. Egre & N. Klinedinst (Eds.), *Vagueness and language use* (pp. 121–145). Palgrave Studies in Pragmatics, Language and Cognition.

Solt, S. (2014). An alternative theory of imprecision. In T. Snider, S. D'Antonio & M. Wiegand (Eds.), *Semantics and linguistic theory (SALT)* (Vol. 24, pp. 514–533). Washington DC: LSA and CLC Publications.

Evaluating Semantic Co-creation by Using a Marker as a Linguistic Constraint Tool in Shared Cognitive Representation Models

Stefan Schneider and **Andreas Nürnberger**

1 Introduction

Information overload has become one of the most critical challenges in humans history. It has been shown that speech, writing, math, science, computing and the Internet are based on independent languages, which together form an evolutionary chain of languages as response to information overload (Logan 2006). Recent technological developments like internet of things (Färber et al. 2020) or cognitive augmented reality (Chi 2009) make clear that this chain continues to advance. Researchers must therefore ask themselves which approaches are suitable to cope with the new levels of complexity.

Cognitive representation models are a key element of the presented evolution, not only with the individual, but also where they can be used as artefacts in conversation. A cognitive representation model can be understood as an abstract model, from which an individual can infer the relationship of objects to one another his environment (Kaplan et al. 2017). Typically, the objects are related based on their properties. Considering e.g. a scale from "tiny" to "big", a "needle" would very strongly relate to "tiny", whereas "mountain" relates more to the "big" property. Using cognitive representation models in a collaborative manner can improve situations in which someone tackles a problem solving task, such as human-robot-interaction (Spranger 2016), within a complex environment. In such situations, people communicate successfully if they come to the conclusion that they are talking about the same things and their cognitive representations converge (Brennan 2005). Such a convergence is known as semantic co-creation (Gergen 2009). Evaluating this phenomenon becomes dif-

S. Schneider (✉) · A. Nürnberger
Data and Knowledge Engineering Group, Faculty of Computer Science,
Otto-von-Guericke-University Magdeburg, 39106 Magdeburg, Germany
e-mail: stefan.schneider@sschneider.de
URL: http://www.dke-research.de

A. Nürnberger
e-mail: andreas.nuernberger@ovgu.de

ficult under realistic conditions, while many factors (like technical communication problems) can bias the given observation (Kraut et al. 2002). Simple collaborative identification tasks (also named referring expression tasks) enable the evaluation of shared cognitive representation models under controlled laboratory settings and allow to observe the progress on semantic co-creation moment-by-moment (Brennan 2005). Findings on how to reach a state of semantic co-creation more easily are helpful in developing adaptive systems that make the complexities of the environment easier to use.

Previous work on referring expression tasks have got a tradition in evaluating collaborations which have a shared space (Kraut et al. 2002; Brennan et al. 2008; Neider et al. 2010; Müller et al. 2013; Hanrieder 2017) as well as a shared cognitive representation model (Brennan 2005; Keilmann et al. 2017). A basic example could be two people who try to find a particular street of a city together by sharing a geographical map. One person who is familiar with the location of the street could explain the route to this target by referring to places in relation to the target street which both participants are familiar with.

In one specific case researchers wondered about the benefit of using markers within these shared artefacts to improve the coordination behaviour based on visual evidence. The question arises if a shared marker can support the participants to achieve a state of semantic co-creation based on a shared cognitive representation model. A shared marker can be anything (like shared gaze (Brennan et al. 2008), shared mouse (Müller et al. 2013) or shared location (Keilmann et al. 2017)), which can be used in shared space or cognitive representation model as a spatial indicator (Müller et al. 2013). Results in this field state that shared markers are in general a beneficial tool (Brennan 2005; Brennan et al. 2008; Hanna and Brennan 2007; Neider et al. 2010; Müller et al. 2013). This becomes obvious when we reconsider the example about human behaviour regarding travelling. In the simplest case, a marker could be a finger of a participant moving across the map in order to explain or support a description non-verbally. If participant A says: "Drive straight through the small street until the next crossing is coming!" Participant B moves his finger along the road, in a manner he has understood the utterance of participant A. Once participant B has moved in a sufficient manner participant A will continue his description, e.g. "Ok! From this crossing, then turn left again." In a case of misconception, participant A would e.g. say to participant B: "No! I meant another street." The finger as some kind of marker applies the participant's given conception onto a map, which indicates to the other participant which aspects were comprehended correctly. Using such a marker (pointing to portions of a map) in addition to a cognitive representation model (a map) appears to be successful in solving a collaborative language task. All participants are informed promptly, using the model, about what has currently been understood (Kraut et al. 2002).

Despite the obvious benefit of using a shared marker, the current research results are not as clear as it might be expected. Specially, the problem is that even any study enforces the usage of a shared marker, while the task durations are very short. For example, Brennan observed task durations between 10 and 20 s (Brennan 2005). Such short durations let us infer that no real team interaction occurred, and then

using a marker is no benefit but only a requirement to finish the task. Based on this contribution, we want to assess the benefit of a shared marker when it is optionally used in comparison to an increased decision complexity. Decision complexity is a user constructed criterion based on the number of alternatives available (Payne 1976). It has been shown, that configuring structural properties of a shared space (Kraut et al. 2002) or cognitive representation model (Keilmann et al. 2017) can influence the perception and even the communicative success. If we use a shared marker optionally, then it can be understood in the manner of a linguistic constraint tool. Such tools relate in some way to linguistic constraints, which cover by symbols the application of dynamic constraints in language use. We hypothesize that the value curve of a linguistic constraint tool based on a given cognitive complexity determines if it is useful or not in given situation of collaboration. In this study we will show that if a shared marker becomes optional under less decision complexity, then it becomes too expensive to use them.

This contribution is structured in the following manner: Collaborative task settings using conversation can be explained by using the contribution model (Sect. 2). Based on the contribution model, semantic co-creation happens when the grounding criterion is reached. There are forces that influence the nature of contributions within the discourse, named linguistic influences. Linguistic influences on the grounding criterion have yet to be investigated in research. Hence, we explain in detail the concept of linguistic constraints. Here, we describe how linguistic constraint tools—represented for example by using a cultural artefact—can influence collaborative task performance (Sect. 3). We explain a new setting, where a marker is applied as a linguistic constraint tool based on a given cognitive representation model (Sect. 4). Based on our theoretical considerations we specify a research design based on a marker and complexity condition (Sect. 5). This design is embedded into a geographic map as the most intuitive cognitive representation model. Furthermore, we describe our collaborative task of identifying a target location, to evaluate the role of a shared marker in addition to a cognitive representation model. The described setting is a very common task, which enables participants to participate without any prior briefing necessary. The principle of least collaborative effort becomes continuously visible to the team by implementing delay discounting decision problem into the reward system. The setup becomes complete through the description of testing conditions in the manner of applied communication media and representation model constraints. Based on this specification, we describe the applied procedure in detail (Sect. 6). Central to our procedure is an implemented chat-tool integrated into a shared geographic map. While three participants are meant to solve the task at three working stations without any moderation, the tool provisions step by step the testing conditions and monitors the progress of a game round. The results show that we cannot observe the characteristics of team focused interaction (Sect. 7). With the first level of decision complexity, no real team interaction occurred. Based on a second level decision complexity, more intense team interaction occurred, but the marker condition achieves in general a disadvantage. Theoretically it is assumed that if participants collaborate they will be most successful if the discussion is constraint in some fashion (team focused interaction hypothesis) (Sect. 8). While our research

design tries to confirm the team focused interaction hypothesis, the results contradict their assumptions. From our point of view, decision complexity seems an important control parameter, which has not been covered with the given team focused interaction hypothesis.

2   Contribution Model of Conversation

Tomasello (2014) states that one basic advantage humanity has is the capability and motivation to collaborate and to help each other. Some human activities are only possible when multiple people are able to coordinate in a highly complex way (taking for example "playing a duet on a piano"). The contribution model of conversation contains a basic approach to explain how long the participants are interested in collaboration, or not. The model explains the coordinative behaviour of participants through internal economic forces. This is based on the assumption that if people participate in a conversation they act in a collaborative manner. Here, we summarise the basis of this model, which is also explained in other contributions by Clark and Bangerter (2007), Clark and Brennan (1991) and Clark and Schaefer (1989) (see also Fig. 1).


Fig. 1   Previous contributions: Team focused interaction for shared cognitive representation models having a marker

Coordination, common ground, semantic co-creation, contribution: In collaboration through conversation people face the problem of coordination, which is implemented by using contributions in participatory acts (Clark and Schaefer 1989). In participatory acts, people act together, which requires them to synchronize in terms of content and timing. Taking musicians as an example, playing a duet on a piano. Both musicians have to confirm, which duet they would like to play (coordination of content) and while they are playing they have to synchronize their entrances and exits (temporal coordination). To enable people to coordinate in conversation efficiently, they need to build a form of common ground. Common ground can be understood as an invisible form of cognitive representation, which all participants accept. In communication, common ground cannot be properly updated without a process. The question is how to evolve an individual idea or conception of something into a form of community-wide accepted semantic co-creation, that is manifested by using a state of common ground (Gergen 2009). Plain and simple, semantic co-creation is given if all task-related participants have got a sufficient idea of how to solve a problem in a collaborative manner, successfully (Raczaszek-Leonardi and Kelso 2008). Participants achieve successful coordination when they reach two degrees of semantic co-creation: grounding and identification (Clark and Wilkes-Gibbs 1986). In identification, participant one tries to get another participant to pick out an entity by using a particular description. Identification happens as soon as the pick out behaviour of the second participant is visible to participant one. In contrast, grounding happens when both participants think that they have identified the correct entity. This means the entity has already been added to the participants' common ground. To put it in a nutshell, the required description to reach a common ground (content specification) and semantic co-creation form a unit of discourse, so called contribution (Clark and Schaefer 1989).

Reaching the grounding criterion using the least collaborative effort: In order to evaluate their conversation, the participants have to set a grounding criterion. A criterion that participant A was successful in describing, could be given if participant B takes the correct object. The grounding criterion is achieved, if a certain amount of effort was provided by the participants to reach a sufficient degree of confidence in the success of a communicative act with a specific purpose. In context of a given communication purpose, the grounding criterion is achieved when all participants believe that they have sufficiently understood (Clark and Schaefer 1989). The participants try to reach the grounding criterion with the least collaborative effort. They are motivated to minimize their amount of work by providing dialog contributions that are as efficient as possible. The concept of least effort was classically described by Grice's saying for quantity and manner (Grice et al. 1975). Grice's saying for quantity states: "Make your contribution as informative as is required; do not make your contribution more informative than is required". While for manner: "Be brief (avoid unnecessary prolixity)." If a contribution follows these sayings it is considered proper, that means the participants believe a contribution will be readily and fully understood by their addressees (Clark and Brennan 1991). Nevertheless, the principle of least effort does not make any exceptions for time pressure, errors or grounding (Clark and Wilkes-Gibbs 1986). For example, when under pressure participants may not be able to plan

well-formulated brief statements and in such cases the model of least effort fails. To overcome these problems the principle of least collaborative effort was formulated by Clark and Wilkes-Gibbs (1986) as follows: "In conversation, the participants try to minimize their collaborative effort—the work that both do from the initiation of each contribution to its mutual acceptance." In participatory acts the participants have to reach the grounding criterion, while minimizing their effort, this is characteristic of conversation in general.

Contributions as a historical process: Ongoing contributions of a discourse have to be considered in historical fashion (Clark et al. 2007). In a classical referential communication task by Krauss and Weinheimer (Krauss and Weinheimer 1964), a participant has to describe his partner which of the four presented abstract figures needs to be selected. To identify the correct figure, each team requires a number of descriptions and related feedbacks. The results of this experiments confirm that ongoing user interaction leads to coordination as a historical process; meanwhile the common ground is constantly emerging. Therefore, as interaction is continuing, descriptions become even shorter ((Krauss and Weinheimer 1964): e.g. (1) "the upside-down martini glass in a wire stand", (2) "the inverted martini glass", (3) "the martini glass" and (4) "the martini") and the number of required turns decreases over time (Clark and Wilkes-Gibbs 1986). In order to make descriptions gradually become more efficient, a form of functioning interaction requires some kind of working user interaction. The average length of descriptions only drops if participants can give direct feedback.

3 The Influence of Linguistic Constraint Tools on Reaching the Grounding Criterion

Bias on reaching the grounding criterion: The previous section demonstrated that reaching the grounding criterion is fundamental for communicative success. For this reason, it is important to understand how reaching the grounding criterion can be influenced. Answering this question is about looking for "tools", which are used to let semantic co-creation happen. Modifying the performance of this tools influences the success on reaching the grounding criterion. Three types of tools are required for reaching the grounding criterion: These are signs, practices and a communication channel. We follow the notion of Löbler (2010), while a sign is "everything which is perceivable, everything we become aware through the senses."; and further a practice "coordinate ways of doing and sayings." Further he noted that: "Practices are implicitly behind all forms of explicit coordination, they coordinate implicitly, and we can become aware of them by the ways we do or say things." If signs and practices wants to be applied, then a communication channel has to be used to overcome a spatial distance in the shared environment. Here we follow the basic channel notion of Shannon's sender-receiver-model (Shannon 1948): "The channel is merely the medium used to transmit the signal from transmitter to receiver."

It has been already pointed out, that practices, as well as the resources available within a communication medium, are termed critical factors (Clark and Brennan 1991). Reaching the grounding criterion can easily translated as what needs to be understood with a given purpose (Grice et al. 1975). This criterion changes through the application of a specific content practice suitable for a given purpose (Clark and Wilkes-Gibbs 1986). For example participants has to identify objects, then a conversation focus on them and their identities. The applied content practice has to ensure that the objects can be identified securely and quickly. Based indicative gestures as an exemplatory practice, an object identified if a speaker refers to an object nearby and the addressee can identify them by pointing, looking or touching. A communication medium (like e-mail or fax) got also an effect on reaching the grounding criterion, while their fulfilling of communication channel constraints differs (Clark and Wilkes-Gibbs 1986). There is a set of costs (e.g. formulation costs or understanding costs) that can quantify these constraints from different perspectives. Nevertheless the influence of signs has been not respected, yet.

Constraints: In this study we follow the idea of linguistic constraints by Pattee (1997) reformulated by Rascazek-Leonardi and Kelso (2008). A fundamental premise of Pattee's theory of living organisms states that there is an interelation between measurement and control. Here, control is about producing a desirable behaviour in a physical system by imposing additional forces or constraints. These constraints are not fixed, but are applied and adapted based on the demands of the environment. They are applied dynamically, following the purpose of a coordinated action.

Linguistic constraints: The described notion of constraints is limited to a specific moment and place. In addition to control, measurement is a symbolic result of the dynamic process. While constraints in a moment of control was limited to a certain point in time and space, the emerging linguistic constraints in the momement of measurement are not fixed. Linguistic constraints—instantiated through symbols— encode stable patterns of dynamic variables that are relevant to control something between an individual and some environment. The human's task of measurement is to choose a relevant pattern and ascribing a symbol to it. Together, linguistic constraints applied in measurement and control can only be understood in a given situation and context of a given space and time they are applied. They are covering the history of constraint application in language use based on multiple timescales.

Linguistic constraint tools: Bringing linguistic constraints into practice we have to notice that typically they are embedded. Hence, Löbler pointed out that "signs render services in helping to find what we are looking for (Löbler 2010)." For us it follows that linguistic constraint tools instantiate services in relation to linguistic constraints. These tools e.g. discussion, cultural artifacts, cognitive representation models or marker can be valuable to achieve a state of semantic co-creation more easily.

4 Using a Marker in Shared Cognitive Representation Models as a Linguistic Constraint Tool

In the last section we introduced the concept of linguistic constraint tool. The question remains open how linguistic constraint tools can help in achieving semantic co-creation based on shared cognitive representation models. In this section we present notion of team focused interaction and summarize previous findings on using a marker in cognitive representation models as an example of a linguistic constraint tool.

Origins of team focused interaction: Team focused interaction describes an approach to identify the correct target in situations of high decision complexity (e.g. identify a target from many). The hypothesis was introduced by Zubek et al. (2016). They evaluated the constraining role of cultural artefacts on the performance of a collaborative language task within a real world setting. The authors used a wine-identification task. Participants were separated into pairs and single probands. In contrast with single probands, pairs can talk freely to each other. Namely, they tried to identify wines, based on their shared tasting experience. Every pair has to talk about smell experience of wine, so it can be assumed that based on the same purpose, there might be similar practices applied. From external, the conditions of communication for pairs were the same. They can talk freely to each other, as long as they want in place they shared physically.

The cultural artefact was a wine tasting card that contains 21 items including a category and their available attributes in the field of taste, smell and general characteristics of wine. For example there was category "Alcohol" and an attribute "Light". In team interaction, the participants can use these taxonomy to describe their tasting experience. Their experiment was designed to evaluate the identification performance based on two conditions: the use of a wine tasting card and whether one participant uses this card or whether two participants use the card and interact freely.

The results showed that interacting pairs were better in identifying the correct wine than an individual wine taster. With the help of a wine tasting card the accuracy of individual participants did not improved significantly. The best performance was achieved, when a pair of wine tasters used a wine tasting card. Pairs using a card had more consistent vocabulary, than those without. The more consistent their vocabulary, the more they were successful in identifying the correct wine. In addition, participant pairs using a wine tasting card had a lower variance in their identification of wines compared to participant pairs without such cards. The lower variance relates to the usefulness of wine categories within the wine tasting card. These linguistic categories likely function as linguistic constraints by focusing the communication, making wine identification more reliable and precise. Together, a linguistic constraint tool can stimulate team interaction towards more focused communication, we name that idea the fundamental premise of team focused interaction.

Nevertheless, this study did not use a common cognitive representation model and even shared marker was not present within them. In both cases, the cognitive representation model was created by each participant, separately in their minds.

There are a couple of referring expression tasks, which cover our research interest as combination of having a shared marker, by using a cognitive representation model and with respect to decision complexity (see also Fig. 1).

Kraut et al. (2002): This study by Kraut and colleagues investigates the role of decision complexity on the communicative success. Decision complexity controlled by the puzzle difficulty (easy: non-overlapping, complex: overlapping elements) and the color drift (easy: static colors, complex: changing colors). These two measures are contrasts with respect to having a shared space, delayed (3 s delay) or not delayed. For evaluation purposes, a team of participants has to arrange a puzzle in the correct order. One participant explains to another how he has to re-order the elements. The results show that teams become faster in solving the task when the puzzle uses non-overlapping elements and static colors were present. Especially, changing colors become a problem to the participants if the screens are not shared immediately. Timing the utterances in discussion moment-by-moment becomes even more complex, because utterances meant to achieve semantic co-creation are biased. In this respect, it becomes obvious that achieving semantic co-creation is influenced based on the perceived decision complexity.

Brennan (2005): Brennan wants to observe the convergence of semantic co-creation moment-by-moment. She is especially interested in how the fact of having a shared marker as a visual indication can influence this process. The research design compares a factor of visual evidence (having a shared marker or not) in contrast with map familiarity (being familiar with a map or not). Based on a geographical map, a participant has to get to an unknown target location, which is visible and explained by a second participant. The participant's mouse movement gives evidence of what has been understood based on the speech (description) of the first participant. Brennan defines several time stages of comprehension towards semantic co-creation: start, first move, close to the target, reliably understood but not at the target, pause and identified. The results show that participants who have no shared marker not only require more time to finish a task, they also require a lot more words to get from a reliable to a final state, and they require a lot of time at the pause stage to get a collaborative acknowledgement. To sum up, a shared marker acts as a visual linguistic constraint tool that simplifies the process of semantic co-creation. Participants become more efficient when a shared marker is present. A second observation shows that solving decision conflicts becomes much more difficult without a shared marker. The final acceptance phase requires much more time than with the presence of a shared marker. Regarding this, decision conflicts can be present as a natural part of semantic co-creation.

Hanna and Brennan (2007): The authors ask if coordination by shared gaze can outperform speech. In the previous study, Brennan has designed shared gaze as something that gives visual evidence. In this investigation shared gaze is introduced as a new type of linguistic constraint tool in addition to discussion based on speech. The role of shared gaze in discussion is evaluated through structural task properties (the orientation of available elements and the distance of a competitor in relation to the target) that control the perceived decision complexity. The task requires to identify the correct target element from a set of forms that are presented in front

of the participant. A second participant knows the correct element and describe this target from another side of the table. Both participants are recorded by eye-tracking and voice recording. For evaluation purposes both recordings are integrated into one stream. The results show that eye gaze produced by a speaker can be used by an addressee to resolve a temporary ambiguity, and it can be used early. Shared gaze outperforms speech because its orientation becomes even faster. This observation let us conclude that there is a competition between shared gaze and discussion, which is won by shared gaze.

Nevertheless, the results hold true only up to a certain level of decision complexity. Shared gaze outperforms speech only in cases with no similar objects (no-competitor condition) or far distant similar objects (far competitor condition) in the possible answer set. In the case when there are similar objects close to each other (near competitor condition), no advantage of shared gaze in contrast with speech can be observed. That means a complementarity of multiple linguistic constraint tools as already described for multiple timescales is needed in cases of high decision complexity. In this respect the complementarity happens in time (multiple timescales) and space (multiple tools) as well.

Brennan et al. (2008): In a further study Brennan is interested in collaborative search scenarios in which both of the two participants are not aware about the target. The collaborative search was evaluated under a shared-gaze, shared-speech, shared-gaze and shared-speech or no sharing conditions. In their study participants had to identify a possibly present O within a set of Qs (O-in-Qs search task). Where participants searched together without sharing anything, accuracy was very low, in fact accuracy was lower than where individuals searched alone. Under shared-gaze conditions the best results, in terms of search duration and accuracy, were achieved. A longer search was observed in teams under conditions of shared-speech or shared-speech and shared-gaze. That observation that shared-gaze outperforms shared-gaze and shared-voice is obvious. The given task is clear to participants without any further negotiation to finish them successfully. If no semantics needs to emerge from collaboration, then no semantic co-creation is required. Even so, team focused interaction is not only about collaborating users having a linguistic constraint tool, it contains also a successful coordination based on semantic co-creation. That means having a shared marker (e.g. shared gaze) can outperform a combination of shared marker and discussion, but only in cases where semantics does not have to emerge.

Neider et al. (2010): Based on the previous study of Brennan et al. (2008), Neider et al. are interested in collaborative search scenarios with two participants working as novices. In contrast, a scenario is studied where consensus between the participants is required. Hence, the study requires that both participants together have to identify the correct target to finish them in time-critical manner. The study design compares shared gaze only, speech only, and shared gaze plus speech. In addition, a no communication condition is applied. In a sniper task, a virtual environment is used to identify the correct sniper target together. The results show that shared gaze together with discussion outperforms shared gaze alone. This observation confirms the previous assumption that semantic co-creation needs to be required to unfold the benefit of team focused interaction.

Further, Neider et al. evaluated that in these cases the principle of least collaborative effort becomes true. Having shared-gaze in contrast with speech only condition, the first participant was faster in identifying the target location because its location doesn't need to be described in detail. If the situation becomes clear to the second participant the first participant only has to note that the second has to go to the target and the task was solved successfully. In a not so clear situation, monitoring gaze behavior as well as more scenic descriptions slow down the consensus phase. Two scenarios with different costs on consensus side are observed. Only when necessary the participants dive into a more detailed discussion. Based on the principle of least collaborative effort such a behaviour can be expected.

Müller et al. (2013): Investigates the role of discussion context and the question if shared mouse can approximate shared-gaze behaviour. They applied a puzzle arrangement task and distinguished between sharing a common gaze, a common gaze and speech, sharing a mouse and speech or speech only. A participant had to arrange the correct order of a puzzle from a set of randomly organized puzzle elements based on the description of a second participant. In addition to the different sharing conditions, one group of participants had to strictly follow particular instructions (low autonomy), while another group could rearrange the puzzle freely (high autonomy). The level of autonomy showed the strongest effect for all communication conditions. Low autonomy conditions resulted in better task-performance, based on lower error rates, independent of the communication conditions. This shows that in a given task-context, more specifically having high or low autonomy, acts as a constraint in communication. The results of Mueller and colleagues underlines the observation that interacting-pairs perform best when they are restricted by some form of linguistic constraint tool. In contrast to Zubek et al. (2016), Muellers study did not require a predefined taxonomy—like a winetasting card—in order to enforce a specific behaviour of the participants. Furthermore, having understood autonomy as a discussion rule tool is no shared marker. Benefits and costs are quite different. Based on such a discussion rule, e.g. continuous monitoring as described by Neider (2010) is not needed. The results also show that in cases of low autonomy, the shared gaze and discussion condition perform even better than discussion condition alone. That means several linguistic constraint tools can be used at the same time. Discussion, shared marker, shared taxonomy or even conversation rules are only some examples of such tools.

By comparing shared mouse and shared gaze, it was additionally observed that shared mouse becomes a good approximation of shared gaze to the given task. Solution times were within the same range and error rates were only higher for gaze than for mouse transfer when the former was used without a speech channel. Considering shared mouse a visual indication requires much more time then shared gaze, but in contrast shared gaze provides much more marker data which has to be interpreted by the participant in a sufficient way. Summarizing, linguistic constraint tools are more or less suitable based on the current purpose of coordination.

Keilmann et al. (2017)[1]: In our terms, Keilmann and colleagues try to evaluate the team focused interaction hypothesis, where a cognitive representation model becomes a linguistic constraint tool. They compared a partly visible and a completely visible labyrinth to one another, either with an individual or with participant pairs. The study examined having a labyrinth as an example of a cognitive representation model or not and whether participants work in a team or alone. If the labyrinth was shared, both participants could see the complete map. In contrast, if it was not shared only a specific subarea of the labyrinth was visible to each participant, individually. The current position of a participant as shared marker was only shared if a participant was located in the visual field of the other. As a third factor, the perceived decision complexity was controlled based on the number of intersections present in the labyrinth. Communication between the participants in collaboration was allowed via headphones. As fast as possible, the participants have to search the complete labyrinth to get all pickup items. The results show that collaborating participants in contrast to an individual searcher are faster and require less trajectory lengths, even though they get higher error-rates in picking up correct items. Hence, team interaction is more expensive than searching alone, but together pairs can achieve a better identification performance. These observations refine the idea of Zubek et al. (2016) that team interaction improves the identification accuracy but requires more communication costs to achieve a coordinated behavior. If the cognitive representation model was shared teams generally outperform single participants. We consider the labyrinth as an example of a cognitive representation model, which is another linguistic constraint tool. Using a shared cognitive representation model collaboratively enables interactions to be more focused.

Hanriede (2017): Hanrieder transforms the Keilmann's stimuli from a top-down into a within-environment view. The participants in the role of firefighters search a floor for casualties as fast as possible. The given task was applied with two cooperation modes (either as individuals or pairs) and several levels of labyrinth complexity (8, 11, 14, 17, or 20 intersections per environment). Hanrieder's setup comes very close to Keilmann's, but in contrast it provides no shared cognitive representation model and it prohibits communication between the participants, who work in teams. Beside the mode of cooperation (individual vs. team collaboration), the decision complexity is controlled based on the number of intersections. It can be shown that the number of intersections have got a negative impact on the task performance. The participants be it individuals or pairs, needed less time to finish, travelled shorter distances and got less error-rates (missed less pickup items). More detailed than Kraut, 2002, it was possible to observe that an increasing decision complexity leads to higher costs and error rates. Keilmann's observation that groups in contrast with individuals are more expensive, while the error-rates are on a lower level which can be confirmed also in a virtual environment.

---

[1]Note: The contribution by Keilmann et al. does not seem to be available for public use. Hence, our description is based on explanations by Hanrieder (2017).

Additionally, the degree in division of labour is measured by the self-overlap of the participants (number of locations at the labyrinth which have been visited more than once) in comparison between individual and pairs. The results show that individual and pairs achieve less overlap for more complex environments. Comparing pairs to individuals, pairs achieve much less self-overlap than individuals. With regard to team focused interaction, this insight is quite interesting because division of labour can be improved even if the participants cannot communicate and have no shared tool working as a linguistic constraint. This observation does not contradict the fundamental premise of team focused interaction. Division of labour seems to be a fundamental practice which is enhanced by team focused interaction.

Prediction: In the following, we want to predict the effect of a shared marker if it is used on top of a cognitive representation model. Such a marker is a very flexible user-driven linguistic constraint tool, which is embedded into a shared cognitive representation model. A marker used in addition to a shared cognitive representation model limits the decision space. By using a marker each participant can time their words and actions better. At any moment of content specification and semantic co-creation, the participants exchange evidence via the shared cognitive representation model, whether the grounding criterion is fulfilled or not. In addition, a marker position informs all participants how far they are from reaching the grounding criterion. If the marker was not moved, then only the given cognitive representation model can be used to limit the decision space.

Using a shared marker is a very common setting in the presented referring expression studies. If such a shared marker is present then each case using them is enforced. The participants have to move the shared marker to fullfil the task successfully. Such an "enforced move requirement" is a problem because it prevents a fair competition between the shared marker and discussion as two ongoing linguistic contraint tools. We are interested in the question if the fundamental premise of team focused interaction becomes true, even if shared marker usage becomes optionally. If we compare experiment durations of previous studies, than we can observe that study durations can be grouped in two categories. The first category experiments are those having a total duration up to 20 s, while the second category observes durations up to 140 s. If we are think about the nature of discussion, then we think that coordination in first category of tasks is very straight-forward. Neider et al. (2017) named that phenomenon one feedback based on description. From our point that means that there is a communication channel but no real team interaction occurs. Such a behaviour can be explained because of a very low perceived decision complexity. Hence we predict that if decision complexity becomes very low than no real team interaction occurs. With second higher level of perceived decision complexity the fundamental premise team focused interaction should become true. Having shared marker should provide an advantage to achieve a good identification performance (Table 1).

Table 1 Comparing previous studies based on the observed task duration

Total durations (approx. mean values)		
Contribution	MIN	MAX
Kraut et al. (2002)	40s	120s
Brennan (2005)	10s	20s
Hanna and Brennan (2007)	n.a.	n.a.
Brennan et al. (2008)	30s	120s
Neider et al. (2010)	8s	12s
Müller et al. (2013)	50s	140s
Keilmann et al. (2017)	n.a.	n.a.
Hanrieder (2017)	50s	120s

5 Setup

The team focused interaction hypothesis claims that using a linguistic constraint tool influences the success of reaching the grounding criterion. In our study we are specifically interested in using a cognitive representation model as a shared artefact, with an additional but optional shared marker available. In collaborative manner the team can solve the given task only based on linguistic features through discussion. It is up to the team to use a shared marker as visual support. The limiting power of a marker will be evaluated by comparing groups with and without a shared marker. The complexity of the cognitive representation model appears to strongly impact the performance of such markers. Hence, we implemented marker conditions with two degrees of complexity.

Which cognitive representation model is suitable for the given evaluation purposes? In our study, a cognitive representation model is present as a shared artefact. There are several forms of cognitive representation models (such as the conceptual space (Gärdenfors 2004), the biplot (Gower et al. 2011) or the associative semantic network (Collins and Loftus 1975)), which differ in their representation (e.g. spatial vs. graph representation) and in their dimensionality and the number of entities they consist of. However, we only applied the geographic map (Monmonier 2018) as a very intuitive example of a cognitive representation model. For our study we required participants to understand the model without any prior learning effort, as such we selected the geographical map as the most widely accepted model. Geographical maps represent complex structures based on standardized criteria (typically distances). A map can be specified as 2-dimensional ($l \times l$ e.g. a map of Germany) or 3-dimensional length space $l \times l \times l$ e.g. an orbit map of our galaxies) By using a globally standardized metric to describe the orientation between very many entities within a space (e.g. all cities in a country), it becomes possible that a large society of people can coordinate within this shared space. For example, millions of deliveries are shipped about the whole world every day only based on one standardized geographic world map. These characteristics make a geographical map very useful where

a large group of people want to coordinate in a shared space (Monmonier 2018). We ensure our model preference by asking the user about the tool familiarity of some other promising cognitive representation models. We measure tool familiarity as the degree model usage within the daily life.

What form should the referring expression task take? We want setup a remote referring expression task in a shared environment, while such a task allows to observe if-based on a given referring expression an intended referent can be picked out (Clark et al. 2007). Such shared environment tasks are possible in two settings: First, the expert-novice setting (e.g. Müller et al. 2013; Brennan 2005), here one participant is familiar with the target (expert), while another participant who is not familiar with it (novice), has to identify it. Second, the novice-novice setting (e.g. Brennan et al. 2008; Zubek et al. 2016) is about identifying a target, while the target is unknown to both of the participants (both act as novices). To evaluate the impact of shared markers on reaching semantic co-creation, it is required that shared markers can play a primary role for coordination purposes between the participants. Hence, we prefer to setup an expert-novice setting, because in such a setting, participants have to collaborate to identify the target. Nevertheless, in novice-novice settings it is possible to search separately (Müller et al. 2013).

The aim of our evaluation is to observe when and how semantic co-creation occurs within a group of participants. We decided to set-up a group of three participants, a describer, an actor and an observer. Under shared marker conditions the marker is visible to all participants, whilst under non-shared conditions the marker is not visible to the participant whose task it is to describe. In such conditions it is possible that the actor (participant carrying out actions) can help the observer (passive participant observing interactions between the other participants) by using the marker.

The task itself should be implementable based on a geographic map as an example of a cognitive representation model. Here, the map task is one example, where one participant needs to explain the route of a map to a second participant (Anderson et al. 1991). The two participants are presented with the same map, the first participant is shown a route marked on the map and is asked to describe this route. The second participant marks down his comprehension of the route, based on the given description. The map task differs from other tasks as the communicative success is measured on a metric scale. Describing a route within a map is a complex task, which requires high intellectual effort of the pair involved. In our study we implement an easier target location task, similar to that used by Brennan (2005). In Brennan's target location task a car icon has to be manoeuvred towards a target location. Only the participant whose task it is to describe can see the target location for the car on their map. The actor tries to find the unknown target location by applying the instructions described to him. The actor can use the shared mouse to relocate the car icon within the geographic map based on the hints given. Sharing mouse-movement is evaluated to uncover the current state of comprehension towards the grounding criterion, continuously. In the study by Brennan (2005) the task is completed once the actor places the car icon very close to the target. Such a setting forces the actor to use the car as an existing marker. However, we want to make the use of the markers optional in

order to evaluate the benefit of using them. Hence, our task ends when one city is selected from a given list of all cities present on the map. In principal, it is possible to finish this task without moving the marker.

How are model considerations implemented within the task? The aim of our task is to make the characteristics of the contribution model visible to all participants at any given moment of interaction. As described previously, the contribution model is implicitly present while participants are interacting in conversation. Nevertheless, there are no defined conditions concerning how brief or detailed each contribution to conversation needs to be in order to be understood. This lack of specificity leads to a bias of incorrect reward assessment by the participants. Applying a time constraint to the collaborative task incorporates time pressure and makes participant contributions briefer (Neider et al. 2010). One disadvantage of such time constraints is that it is harder to interpret how much effort the participants invest in conversation. Hence, we prefer to describe the task the participants have to complete as a collaborative conflict situation of least collaborative effort to reach semantic co-creation.

The "conflict" becomes visible to the participants online through scores assigned to participants' actions based on the delay discounting decision problem (Scherbaum et al. 2016). In delay discounting decision problems, a single participant has got two options of which they have to select the most beneficial one. The first option is named sooner smaller (SS) option, which means the user can get this one very fast but he needs to accept a lower reward value. In contrast, the second option is named later larger (LL) option. This option returns a much greater value to the user, but it is much more difficult to reach it resulting in a long delay. Unlike in the classic single participant approach, multiple participants who are trying to coordinate try to reach the highest degree of value discounting. Based on an initial reward score value each team member has to ensure that their actions reduce the team score as little as possible, while reaching the grounding criterion should be achieved as quickly as possible. In our case, the describer needs to decide whether they want to apply a more detailed description (SS-option) or only slight hints about the location, e.g. using the words "hot" or "cold" (LL-option). If a participant applies his own description, he wants to ensure that he reaches the grounding criterion fast, even though only a small team discount can be achieved. In contrast, if a describer applies only hints, he tries to achieve a larger team discount, while it becomes more difficult for the team to reach the grounding criterion, because such hints are much less informative. In our case the actor and observer need to decide whether to select just a target subset (SS-option) or whether they want to know the exact location (LL-option). Actors or observers who only select a target subset slow down the required grounding criterion, because the correct answer needs only one element within the given subset. This option has the disadvantage that the team score decreases very much. In contrast, if an actor or observer selects a unique correct answer the requirement to reach the grounding criterion is much higher, because only one correct answer needs to be identified. This more delayed option seems charming because it decreases the team score much less. In applying participant action scores in the coordination task, it will be possible to measure the degree to which the SS and LL principles have been

used, interactively. For implementing collaborative delay discounting problems, we decided to implement the text chat tool instead of audio channel communication (e.g. Neider et al. 2010).

What are the test conditions? Test conditions of referring expression tasks can be described using the limitations of communication media (Clark and Brennan 1991) along with the content provided with descriptions or identification skills in order to determine the credibility of other participants (Edwards and Myers 2007). Co-occurrence assumes that the participants are present at the same time. Our task is applied to three participants, who have access to the same shared workspace in different roles. Together with a shared cognitive representation model the participants can communicate by using a shared chat system. Here, they can write and read messages at the same time (simultaneity), they can look at older messages within the chat protocol (reviewability) and read their messages before they submit them into the chat (rereading). In shared space scenarios communication delay becomes an additional critical issue (Kraut et al. 2002). The communication between the nodes based on LAN connection as well as our script performance happens without any perceivable delay. Furthermore, we need to ensure that the task is achieved based only on the geographical map and chat messages available to the participants. No additional communication media should be used (e.g. other messenger services), no other sources of information should be available (e.g. Wikipedia) and no common ground should exist between the participants before starting the task. To guarantee these test conditions, participants worked at prepared working stations, which only offered access to the testing environment. The use of mobile phones was prohibited during the evaluation.

Within the test conditions for the cognitive representation model it is considered the following questions: what do the participants already know about the map (pre-existing background knowledge (Brennan 2005)), how are the cities of the map structured (entity structure) and can the participants use a symbol for a city for communication purposes (symbol entity referencing). Pre-existing background knowledge occurs where participants have some common ground beyond the task, which could help them to complete the task more easily. If our task were to use a map of Germany and the participants were German they could use their background knowledge to identify particular places on a common map faster than if their task involved a map of an area unknown to them all, e.g. Ukraine. Our study, in fact, deploys maps of Ukraine and other countries for which we consider it unlikely that participants will have previous knowledge of. The second factor, entity structure, is about the complexity of coordination within the cognitive representation model. This complexity is indicated by the number of elements and the proportion of reference points from all elements (non-reference points). A reference point is a location with discriminable features, which allow a subject to have a geographical orientation (Sadalla et al. 1980). Having reference points improves orientation in cognitive representation models (Hanrieder 2017). Within most maps there are a set of reference points, in our case popular cities in a country (e.g. Kiev in Ukraine). If a reference point was a target the participants could refer directly to cities based only on their name (e.g. a city which is marked "Kiev"). To identify the correct target the describer could send

this description to the actor, who can identify this place easily, without any further interaction. To avoid such a behaviour, we add a set of randomly chosen less well-known cities, one of which needs to be identified. We introduce these random cities with a symbol instead of a name. This approach prevents town names from being used as descriptions. However, the describer could refer to a town's symbol instead ("go to the town x_1"). To solve this potential problem, we inform all participants that their symbols for the given towns are all different, making such references no longer useful. This should prevent the participants from using the symbols of towns for coordination purposes. It also simplifies the map, because only reference points can be used between the describer and the actor. An increasing number of reference points makes orientation on the map landscape easier. Just as well, identifying a target location become easier if the number of potential decision points is smaller. Low map complexity means there is a large number of reference points and a small number of non-reference points. We defined complexity level 1 (low) as consisting of 5 reference points and 10 non-reference points. Map complexity level 2 (high) comprising 1 reference point and 25 non-reference points.

6   Methods

As each task is limited to a duration of five minutes, each team completes the whole experiment (six trials) within 30 minutes. Both the actor and the observer should identify the target location which means there are two task results per task. In total twelve task results are recorded for each team. Resulting in a total of 156 instances for evaluation.

6.1   Participants

Our task was completed by 13 groups each consisting of three participants, with 6 trial rounds. In total there were 39 participants, of which 17 were female, the average age of participants was 32. All participants had normal vision, or their vision was corrected to normal with glasses or contact lenses. Each of the participants gave informed consent to take part in the study and received a natural gift (a bottle of water or a piece of fruit) after completing the experiment. Each team member of the winning team was given a bouquet of flowers as a gift.

6.2   Apparatus and Stimuli

Stimuli were presented on three laptops simultaneously, each with a normal RGB background on a 14.1-inch screen at a resolution of 1377×768 pixels with 60 Hz

refresh rate. In our evaluation, we use an individually implemented analytics pipeline using a survey tool and a pre-processing tool. The survey tool handled our specified task, we described in the last section. It was implemented using PHP and AngularJS and had an integrated MySQL relational database. The chat environment was based on Socket.io. The presented geographical map was implemented by D3 and TopoJSON in a similar way to the tutorial by Mike Bostock. Additionally, we implemented the Natural Earth dataset of GDAL to create each of the maps which included country polygons and populated cities within each country. The pre-processing tool sets up a database for survey data and transformed this data into a dataframe, which could then be directly evaluated using statistical analysis tools such as IBM SPSS Statistics.

The task user interface: The interface of the team workspace consists of a shared geographical map, a chat system, information on the current reward and remaining time and an area to apply participant interactions, with options such as "add a description" or "select target location" (see also Fig. 3). The describer, whose task it is to describe the target location, views the same geographical map as the other participants, but additionally one of the random cities highlighted. The describer can use two forms of participant interaction. They can describe the target location freely or use pre-defined hint buttons, such as short messages indicating "cold" (far away) or "hot" (close) within the chat system. The maximal message length for communication is 67 characters, comparable to the length of an SMS. Both interactions of the describer are reward related. A short hint ("cold" or "hot") relates the SS-option and
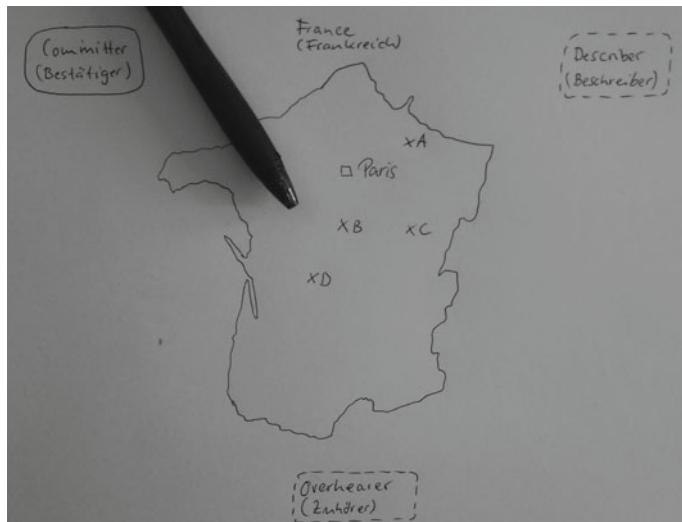
**Fig. 2** *Paper prototype of our map task*: The paper prototype of the map task containing a simplified representation of France including Paris as a reference point (popular city) and four non-reference points (random cities), which are referred to via symbols. In preparation of the three team members, each participant of a team was positioned randomly around the map according to the roles described at the edges of the paper
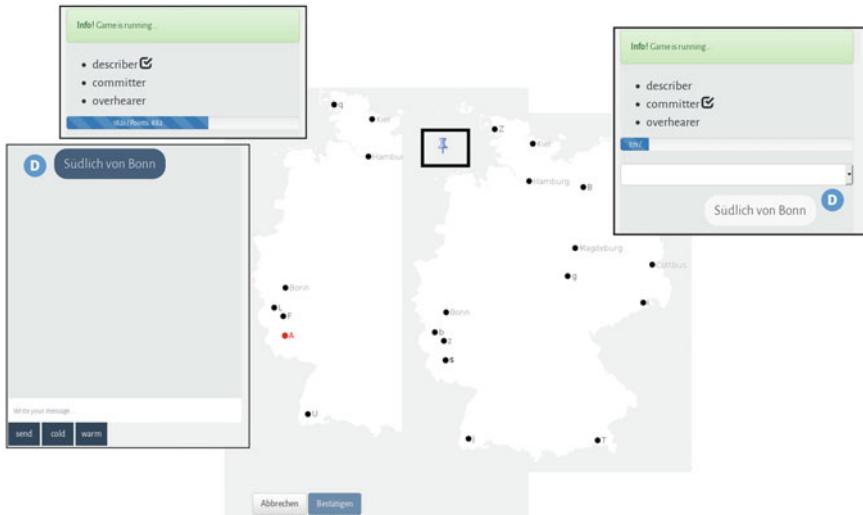
Fig. 3 *Target location task interface*: The user interface has a shared geographical map of a country like Germany containing a set of well-known cities (e.g. Bonn) and hidden cities (marked with symbols e.g. *b*). In the user interface windows on the left, those of the describer, town "A" is marked red, this is the city they have to describe the location of. The user interface windows on the right show that the actor sees the same cities but marked with different symbols. The describer needs to refer to the city of Bonn and indicate with a message that the target location is "south of Bonn". The actor can respond to this message by moving the marker (shown here as a pin above the map) or by replying to the message or selecting a reply option for a predefined answer set

reduces the reward by 10 points, sending a longer text message is the LL-option, which reduces the reward by 50 points.

The concept of collaboration: All participants are directly updated about communication methods used as they can all see the remaining reward amount. The actor can read the describer's messages and thus move the marker towards the potential target location. This marker is visible to all participants but can only be moved by the actor. Via the marker the actor can indicate where they assume the target location to be based on the describer's messages. The actor can also comment on any given explanation of the describer freely, without any costs regarding the reward. The actor has two options to complete the task: (a) by selecting the target location they assume to be correct (LL-option—select 1 city of 10 options in complexity level 1 or 25 options in complexity level 2) or (b) by selecting a subset of target locations, one of which should be the target location (SS-option—i.e. select 1 of 5 options, while each option represents in complexity level 1 two cities or in complexity level 2 five cities). Both of these options are also reward related: the LL-option results in a further reduction of the reward by 50 points, whereas the SS-option results in a 10 point reduction of the reward. In SS-options, cities are clustered based on their proximity by using same-size k-means clustering.

6.3 Procedure

Preparation: Before starting the task, a moderator explained the task and the participants of a team completed an initial test-round together by using a simplified paper prototype (see also Fig. 2). Besides, the participants had to evaluate how familiar they were with popular cognitive representation models. The moderator also informed the participants about the basic notion of these cognitive representation models by means of an example. The three participants carried out the task in separate rooms, each provided with a laptop. Communication among participants was only allowed within the provided chat room. Participants had to deposit their smartphones outside the room and the provided laptops had no internet access. After an introduction to the task by a moderator, each participant was left alone in their respective room with the laptop and the task without any further discussion with the other team members. Each participant had to sign-in to a shared workspace where they were then randomly assigned a role (describer, actor or observer).

The location identification task: With the shared workspace each participant within a team sees the same map of a country. This country map contains a set of reference points (popular cities of a country, e.g. Berlin, München, Hamburg, Frankfurt or Stuttgart for Germany) and a set of non-reference points. Non-reference points are cities which are labelled by a personally chosen random symbol (like a_1 or a_2) The team of participants begins the task together with a time limit of 5 minutes. The aim of the task is for the actor and the observer separately to identify the correct city as efficiently as possible, based on the hints given by the describer. The reward for achieving this task is 1000 points at the start of the game; this reward decreases as time passes and with increasing participant interactions. The team with the highest score after successfully completing the task wins. Unsuccessful teams who do not manage to complete the task end the game with 0 points. Additionally, the task ends earlier if the reward is reduced to 0 points based on the team's amount of participant interactions. After starting the task, the remaining time is displayed in the shared space, along with the current reward, which decreases at a rate of 1 point per second.

Finishing the task: Once the actor has correctly identified the target location by selecting the correct target location, all participants are informed via the chat that the actor has finished, the location is however not visible to the other participants. If the actor has completed the task first, they should help the observer to also find the correct target location. Hence, an actor who has finished stays on in a running game session and can write messages to the describer and move the flag. Giving answers is not possible any longer. Observers are by nature not allowed to interact with the shared space, but they can see everything which is happening. They can use additional user interactions of actor and describer to identify the correct location. Unlike the actor, the observer cannot interact and send chat messages or move the marker. The observer is truly just an observer who can see what the other participants are doing. In addition, she can apply an answer to finish her task. When the observer and the actor select the same target location they are rewarded with the same number of points. If the observer completes the task by correctly identifying the target location,

afterwards they are not able to give hints and can only continue to observe the other participants' behaviour. When both the actor and observer have completed the task, the team is rewarded with the current reward visible to them on their screens.

The marker condition: When only one participant has been able to complete the task the reward amount continues to decrease until the maximal task duration has been reached. The task is carried out under two conditions: either the marker is visible for the describer or not. Under "no marker" condition, the marker is still visible for the actor and the observer (and can be moved by the actor). Therefore, the marker is only helpful for the participants in these two roles.

6.4 Design

Each participant of a team is randomly assigned a defined role (describer *D*, actor *A* or observer *0*). In each role the task is applied either with the use of a marker (*M*) or without a marker *NM*. This combination of three different roles and two different marker conditions results in a total of 6 trials per team. For example, the following trial-order could be applied: (1) $O - NM$; (2) $D - M$; (3) $A - NM$; (4) $A - M$; (5) $O - M$; (6) $D - NM$. Each gameplay consisting of 6 rounds is based on geographical maps of the same complexity level. Whether a gameplay is based on complexity level one or two is assigned randomly.

A new geographical map was generated for each trial, from a set of the 6 countries. Each country contained more than 100 possible cities.[2] For each country only ten cities were selected as candidate reference points (popular cities), the rest were categorised as potential non-reference points (random cities), which were also randomly selected.

7 Results

Our initial focus was the use of the cognitive representation model. To confirm the suitability of the geographical map as a preferential cognitive representation model we asked participants to assess their usage of four cognitive representation model options (a geographical map, a biplot, conceptual space and a semantic network) in their daily life. The results, based on a 7-point likert-scale (from (1) "I don't know what this is" to (7) "I am using it regularly in my daily life") are shown in Fig. 4. The results reveal that the conceptual space and the biplot were the most unknown representation models. 23 of 39 participants didn't know what conceptual space was or had never used it. Similarly, biplots had never been used by 24 of 39

---

[2]Mexico incl. 1190 cities, Norway incl. 417 cities, Philippines incl. 156 cities, Puerto Rico incl. 242 cities, Portugal incl. 286 cities, Ukraine incl. 510 cities.

	I don't know it	I have never used it	I used it only once	I have used it sometimes, already some time ago	I have used it regularly, already some time ago	I am using it not-regularly in my daily life	I am using it regularly in my daily life
Map	0	2	2	4	1	10	21
BiPlot	9	15	6	9	0	0	1
Conceptual Space	7	16	9	4	2	2	0
Semantic Network	3	6	7	16	3	5	0
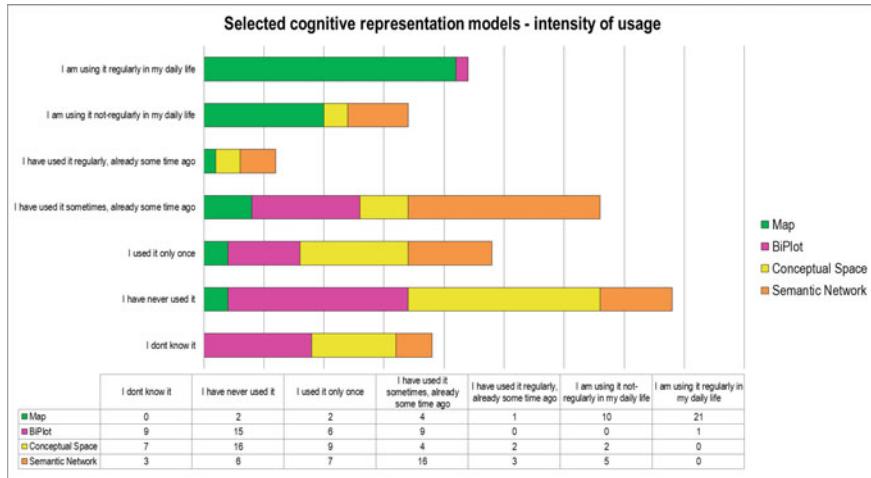
Fig. 4 *Intensity of usage*: 39 Participants evaluate how intensively they use four types of cognitive representation model: a geographical map, biplot, conceptual space and semantic network. The diagram shows how often the selected options (ranging from "I don't know what this is" to "I am using it regularly in my daily life.") were chosen for each cognitive representation model

participants. In contrast, semantic networks were more well-known and used by a larger proportion of participants. Of 39 participants, 24 stated that they used semantic networks, here responses ranged from "I have used it sometimes, but some time ago" to "I use it, but not regularly in my daily life". Nevertheless, the geographical map was evaluated as the most widely used cognitive representation model. 31 of 39 participants confirmed that they used geographical maps, even though not regularly in their daily lives. Based on these results we consider the geographical map as a suitable cognitive representation model for our study purposes, as it can easily be used by a broad range of participants.

We also wanted to evaluate whether the complexity of a geographical map influences the level of interactivity used to complete the task. Based on the principle of least collaborative effort, interaction itself contains the application of linguistic constraint tools which are used more intensely in complex situations. We hypothesised that if the complexity of a cognitive representation model is too low, then no team interaction emerges. To investigate this, we compared the two levels of map complexity. Complexity should influence all levels of interactivity, which describe the nature of a task round. As such we evaluated several indicators of interactivity: how often answers were given, the number of messages the describer sent and how often the actor responded to a message or moved the marker. Table 2 lists these indicators of interactivity. Using the complex geographical map, the describer had to send considerably more long messages (53.3%) and short messages (60.0%) than under simpler map complexity conditions. Where initial descriptions could not narrow down the target location enough, multiple long messages were required. Short messages used

Table 2 The relationship between complexity and interactivity: Several indicators of team interaction are compared with two levels of complexity. Complexity level 1 (2) contains 5 (1) reference points and 10 (25) non-reference points. Complexity level 2 compared with level 1 requires a much higher degree of interactivity

Interactivity indicator	Complexity level...		
	1 (5–10)	2 (1–25)	p-value
Describer			
Sends two or more long messages	29.6%	53.3%	0.006
Gives short feedback (e.g. "hot" or "cold")	12.5%	60.0%	0.000
Actor			
Moves the marker at least three times	11.1%	37.8%	0.000
Sends two or more comments	11.1%	44.4%	0.000
Actor and observer			
Answers more than once (single-answer-option)	0.0%	24.4%	0.000
Answers using the answer set	0.0%	34.4%	0.000

by describers under these conditions tended to be small hints relating to previous interactions with the actor. Under complex map conditions (complexity level 2) the actor used the option of moving the marker more often (37.8% more than under less complex map conditions) and responded to the describer more frequently. An average of 44.4% of all actors gave feedback with at least two comments. Under map complexity level 1 participants selected the correct target city, rather than a subset of target cities. Comparing the map complexity levels there was a significant difference between the number of long messages sent by describers ($p < 0.01$), responses by actors ($p < 0.01$), movements of the marker by actors ($p < 0.01$) under the two map complexity conditions. Overall, there is a significant difference in the degree of team interaction required to complete the task between complexity level 1 and level 2. Figure 5 visualizes these differences based on two session examples. The results show that in contrast to complexity level 2, complexity level 1 requires very little team interaction to solve the task. Only when complexity increases does the level of interaction become more intense.

We also assessed the influence of the marker and the degree of interactivity on communicative success. From our previous observation we conclude that complexity (as a control parameter) influences the nature of interactivity. Hence, we focused on trials with map complexity level 2, where the participants appeared to be under higher pressure to interact. We used our results to evaluate the interactivity hypothesis, initially observed by Zubek (2016): When a pair of participants try to identify a target, they perform better than where one participant attempts this task alone. We reformulate this statement for our study: When a pair of participants interact intensively, their performance (in terms of task completion) is better than when interaction is very limited.

Two degrees of interactivity

Example of a non-interactive session:

1. *describer:* "erste Stadt oben links (westlicher Rand)" ("first city top left (western edge)")
2. *actor:* got the correct answer

Example of an interactive session:

1. *describer:* "süd-östlicher ort" ("south-eastern place")
2. *actor:* "eher nah bei Shiraz oder weiter davon entfernt" ("close to Shiraz or further away")
3. *describer:* "nicht direkt neben shiraz. der entfernteste ort süd-östlich davon" ("not directly next to Shiraz. The furthest place south-east from there")
4. *actor:* got correct answer

Fig. 5 *Two degrees of interaction*: In the first example, the describer sends one message and the actor was able to identify the correct target based solely on this description. Contrastingly in the second example further interaction and refinements are required to select the correct target. It becomes apparent that interaction only emerges when an initial linguistic constraint tool—the message sent by the describer—is not sufficient to reach semantic co-creation (to complete the task)

While we focused on cases where map complexity level 2 was used, interactivity was measured containing only most distributed variables. Most distributed variables are those having the biggest diversity of observed values. In our results, the number of comments made by the actor and the number of movements of the marker were the measures of interaction with the greatest variability. We compared these measurements with indicators of communicative success. The main indicator of communicative success is how many participants successfully completed the task. Here, the answer can be two (the actor and observer), one (either the actor or the observer) or none (neither the actor nor the observer). Further indicators of communicative success are the time taken to complete the task, for the first and second participant in each team to finish. The results for these indicators are summarized in Table 3. Based on these results the hypothesis that participant pairs perform better when interacting more must be rejected. Of all teams requiring 0 comments for finishing a task, 86.5% were successful. Likewise, of all teams that required 2 comments or more, only 40 % were successful. This observation is underlined when we look at the duration when the first participant (be it actor or observer) was able to finish the task. Teams which required no (vs. two or more) comment for finishing a task, were able in 91.7% (vs. 32.5%) to finish a task in fastest half of the first completion times, successfully. Same is true when we look at the number of marker movements. Teams which moved the flag 0 times (vs. 3 or more times) were in 85.0% (vs. 29.4%) under the fastest half of first completion times. Summarizing we can state that the best performing part of a team performs better if they do not interact intensely. Best performing part in this sense means two of three participants having one identifying participant, who was able to complete the task faster. Here, no benefit of interaction on communicative

Table 3 *The relationship between participant interactions and communicative success*: Two indicators of communicative success are compared with two measures of the degree of interaction. It can be summarized that the participants were most successful if they did not interact. However, it has become obvious that interaction is helpful in a way that all participants finish the task successfully

Indicator of communicative success	Number of comments used			Number of marker movements		
Degree of...	0	≥ 2	p-value	0	≥ 3	p-value
All participants successfully completing the task	86.5%	40.0%	0.004	75.0%	44.1%	0.119
Fastest 50% of teams completing the task	12.5%	60.0%	0.000	15.0%	61.8%	0.001
Fastest 50% of participants completing the task	91.7%	32.5%	0.000	85.0%	29.4%	0.000

success can be observed. Nevertheless, the results also make clear that the final completion was improved by a high degree of interaction. Teams which required no (vs. two or more) comment for finishing a task were only at a level of 12.5% (vs. 60.0%) under the fastest half of participants to the full completion of the task. Similarly, of teams which did not (vs. 3 or more) move the marker, 15.0% (vs. 61.8%) were under the fastest half of participants to full completion of the task.

All these observations are significant with a level of at least $p < 0.01$. It should be noted that the number of marker movements did not correlate to the number of participants successfully completing the task.

Our evaluation also considered the impact of a marker as a linguistic constraint tool on communicative success. Adapted from the basic notion by Zubek (Zubek et al. 2016), we hypothesised, that the use of a marker improves team interaction as it focuses communication on critical aspects. More concretely, teams using a marker in intensive interaction should attain the highest communicative success. Based on our results this hypothesis should be rejected. We evaluate sessions with a low and a high level of interactivity in contrast with the marker condition. While the map complexity is only a control variable set from outside, the interactivity level is a phenomenon which is inherently part of team communication. Further the comparison uses the independent variables of investigation in the previous section above. In contrast, we are only interested in the degree of actors who moved the marker or wrote a comment based on a given description. We can observe that having a marker, generally becomes a disadvantage. Of all users who moved a marker once or more often, 78.6% (non-interactive sessions), 70.0% (interactive sessions) were successful in finishing the task while the marker was not visible to the describer. In contrast, when the marker was visible to the describer only 61.5% (non-interactive sessions), 40.0% (interactive sessions) finished the task successfully. Same is true if we look at the first completion time. When the marker was not visible to the describer 91.7% (non-interactive sessions), 85.0% (interactive sessions) were under the fastest half of first completion time. In contrast, when the marker was present to the describer

only 32.5% (non-interactive sessions), 29.4% (interactive sessions) were under this fast half subset. A similar observation occurs when we evaluate those sessions in which the actor applies minimally one comment. Here, we can see a disadvantage, when teams were heavily interactive and using the marker. The numbers of all users that finished with writing at least minimally one comment differs between 75.0% of teams having no marker and not being interactive to 23.1% of teams having a marker and being interactive. Furthermore, 80.0% of all teams using no marker and not being interactive ranked under half of first completion time. In contrast, interactive teams which used a marker and wrote at least one comment, achieved only 30.8% of the first completion times. With reference to marker movement only, we can further observe that interactivity was helpful to finish a task to the slowest identifier (nevertheless if it was the actor or the observer). Interactive teams where the actor moved a marker at least once, were under the fastest half of full completion times in 70.0% (no marker condition) 63.3% (marker condition) of all cases. In contrast, non-interactive teams were not so fast, only 24.1% (no marker condition) 30.8% (marker condition) were under the fastest half of full completion times.

In sessions with intense interaction while the marker was visible to the describer and all other participants, few teams completed the task (23.1%) successfully and also few subteams were under the fastest half of first completion times (30.8%). In contrast, teams with a low degree of interactivity and no marker shared with the describer achieved the best communication success. Under this condition, 75.0% of all users finished the task successfully, while 80.0% were under the fastest half of the first completion time. It can be concluded, that especially having a marker as a linguistic constraint leads to disadvantage in achieving communication success. It needs to be noted, that our first observation doesn't reach the required significance level of $p < 0.05$. Moreover, we have decided to add this observation to our considerations because the level of significance reached is very close to the required significance level and this observation supports our big picture. A further observation of the full completion time based on comments is not under consideration because it is not significant (Table 4).

8 Discussion and Conclusion

The aim of the contribution model of conversation is to reach the grounding criterion with least collaborative effort, which is fundamental to achieve communicative success based on semantic co-creation (Clark and Schaefer 1989). Linguistic constraint tools play a critical role in reaching the grounding criterion within interactions. Hanna and Brennan already mentioned the importance of constraints even outside of discussion (Hanna and Brennan 2007). Nevertheless, the applied constraint model is focused with text-processing and can not explain the adaptive nature of these constraints. The notion of linguistic constraints (Raczaszek-Leonardi and Kelso 2008; Pattee 1997) overcomes this limitation based on clear epistemological roots. Previous

Table 4 *The relationship between interaction and the marker condition on communicative success:* The independent and dependent variables are compared separately, based on the interaction level and the marker condition

Conditions	2 × 2 setting				
Interactivity level	Few interactions		Substantial interaction		
Marker condition	Not-present	Present	Not-present	Present	
Indicators of communicative success					
Degree of...	Actors, that moved the marker once or more often				p-value
All participant successfully completing the task	78.6%	61.5%	70.0%	40.0%	0.051
Fastest 50% of teams completing the task	24.1%	30.8%	70.0%	63.3%	0.003
Fastest 50% of participants completing the task	91.7%	32.5%	85.0%	29.4%	0.000
Degree of...	Actors that wrote 1 or more comments				p-value
All participants successfully completing the task	75.0%	73.3%	71.4%	23.1%	0.000
Fastest 50% of teams completing the task	35.0%	43.3%	71.4%	53.8%	0.171

research has investigated several linguistic constraint tools—like ontology (Zubek et al. 2016), contextual restrictions (Müller et al. 2013), cognitive representation model (Keilmann et al. 2017) or shared marker (Hanna and Brennan 2007). Linguistic constraint tools can help to improve team interaction and make conversation more focused (Zubek et al. 2016). Previous studies have in common that they prevent an optional use of a shared marker. The users have to move a marker to finish the given task successfully. From our point view, such a setting is unfair because it prevents the natural competition between shared marker and discussion as two parallelly applied linguistic constraint tools.

Previously, the advantage of team focused interaction was observed in complex study situations (e.g. identifying the correct wine through smelling and tasting based on a collection of available wines). Some referring expression tasks using a shared marker require a total task duration of less than 20 s. So, we asked if team focused interaction brings a general advantage, independent of the given decision complexity. Our study design compares the availability of a shared marker and two levels of decision complexity. Controlled by the number of cities in a map, we simulated two situations, where the perceived decision complexity differed. As stimuli we used a shared geographical map that contained a number of cities of which only a very small subset could be used as reference points. Furthermore, we reflected the principle of least collaborative effort as a collaborative delay discounting problem (Scherbaum et al. 2016) to the participants. The participants are not only moment-by-moment aware of how far or close they are from a state of semantic co-creation (Brennan 2005), but also the value perspective is visible to the user moment-by-moment.

Continuously the teams can evaluate the major characteristics of the contribution model of conversation. We classified sessions in low and high level of interactivity and compared them with having a marker shared with the describer or not. Based on our setup each team consists of a triad of participants (describer, actor and observer). This team configuration allows us to design a natural way of having a shared marker or not. If a marker is not shared with the describer, it still becomes useful because it is helpful between actor and observer.

We observed significant differences in the degree of team interactivity under different complexity conditions. At the lower complexity level, participants were able to use simple descriptive messages which were sufficient for accurate target identification. However, it is not enough to provide a channel for team interaction. Up to now the team focused interaction hypothesis assumes that if there are two people who can communicate it improves the identification accuracy in general. The observation in this study shows that while team interaction was possible in principle it was not really happening. One description was sent and immediately the identification happens in the next step. This pattern of discussion was already observed by Neider et al. (2010). Nevertheless, it has become obvious that team interaction by discussion itself requires a specific level of perceived decision complexity to make it an advantage. If the decision complexity is too low, real team interaction is not required.

Considerable amounts of interaction were only observed under higher complexity conditions. Nevertheless, we observed that not heavily interacting participants achieve the best performance in accuracy and task duration. That means if perceived decision complexity reaches a specific level it becomes an appropriate tool. This can be interpreted in a way that team interaction is also some kind of linguistic constraint tool, which underlies the principle of least collaborative effort. Still, we could also observe that participants interacted more intensely when they were having problems with identifying the target. Our results have shown that with a given level of complexity, a shared marker becomes useful to solve conflict situations. In similar fashion Brennan et al. (2005) already observed that if a shared marker was not present, situations evolve where in the end an actor is for a long time not far from the target and still does not identify it correctly. From our perspective a conflict situation relates to an increased perceived decision complexity that is why a marker seems to be advantageous.

In general, our results on complexity level two show, that using a shared marker was disadvantageous for the teams. In teams that had a shared marker the describer required more time to complete the task and they were less accurate in identifying the correct target. In contrast, in teams with a low degree of interactivity and no marker shared the describer achieved the best performance. It needs to be noted, that similar to our previous observation, it becomes obvious that the second identifier was faster when the team interacted heavily. From this observation we infer that the marker (as linguistic constraint tool) is not useful for the given level of complexity. Finding this certain level of complexity was not the aim of this study though.

To sum up, these observations confirm the principle of least collaborative effort towards the team focused interaction hypothesis. Team interaction and focused inter-

action are two separate linguistic constraint tools, having different characteristics, and hence need to be evaluated separately. Perceived decision complexity plays a crucial role so that team interaction becomes beneficial. Only when situations become complex team interaction is more and more helpful. Using a linguistic constraint tool (for example a marker) on top of a cognitive representation model requires a even higher level of complexity, while team interaction is already helpful. It might be the case, that if a situation is perceived as complex then in a first step we add teams and let them interact freely. If this is not enough, then we try to get interaction become more focused by using additional linguistic constraint tools.

Our observations are based on three limitations. First, the study was based on the use of a geographical map as a cognitive representation model. It is not clear whether these results can be replicated using other cognitive representation models. The characteristics of other cognitive representation models could also lead to other limitations regarding linguistic constraint tools. Additionally, using a geographical map implies a scenario in which spatial language is required to achieve a state of semantic co-creation (Spranger 2016). In other scenarios, for example searching for a target within text documents other linguistic practices of language use are required. For each different scenario appropriate linguistic constraint tools should be applied. Secondly, only a subset of linguistic constraints was evaluated. If we had only been evaluating the use of a shared marker within a shared space, we could have distinguished between different types of use, e.g. shared-eye-tracking or shared-mouse-tracking (e.g. Müller et al. 2013). A comparison of the suitability of several possible linguistic constraint tools was however not part of this study. Finally, the results show that the team focused interaction does not become true in general. A low decision complexity leads to an outstanding benefit team interaction and team focused interaction as well. The question remains open, if we generally have to reject the team focused interaction hypothesis, when we use shared cognitive representation models, or if a complex enough situation is required so that the hypothesis becomes true. Moreover, research should include the use of more complex geographical maps, containing for example 1,000 to 10,000 random locations to further evaluate the validity of this hypothesis.

Based on our results, we recommend that further studies focus on complexity when evaluating a specific cognitive representation model. The assessment of tools which are appropriate for implementation in search tasks in cognitive representation models should recommend simple filtering tools when the complexity is low. In contrast where complexity is high more sophisticated filtering tools are required to successfully complete search tasks. Using sophisticated filtering tools in complex environments leads to new situations with lower complexity where simple filters are preferred. Simple and sophisticated linguistic constraint tools can comprise any form of linguistic constraint. In our study we focused on investigating the impact of these linguistic constraints on conversation with respect to the aim of reaching the grounding criterion using the least collaborative effort.

References

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., & Sotillo, C. (1991). The hcrc map task corpus. *Language and speech,34*(4)351–366.

Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions. 95–129.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, *106*(3), 1465–1477.

Chi, E. H. (2009). Augmented social cognition: using social web technology to enhance the ability of groups to remember, think, and reason. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data* (pp. 973–984).

Clark, H. H., & Bangerter, A. (2007). Changing ideas about reference. *Experimental Pragmatics*, 25–49.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on socially shared cognition*, 13127–13149.

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

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*(1), 1–39.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407.

Edwards, C., & Myers, S. A. (2007). Perceived instructor credibility as a function of instructor aggressive communication. *Communication Research Reports*, *24*(1), 47–53.

Färber, M., Svetashova, Y., & Harth, A. (2020). Theories of meaning for the internet of things. *Concepts in Action: Representation, Learning and Application*, *24*(1).

Gärdenfors, P. (2004). Conceptual spaces: The geometry of thought. MIT press.

Gergen, K. J. (2009). Realities and relationships: Soundings in social construction. Harvard University Press.

Gower, J. C., Lubbe, S. G., & Roux, N. J. L. (2011). *Understanding biplots*. Wiley.

Grice, H. P., Cole, P., & Morgan, J. (1975). Logic and conversation. *Syntax and semantics*, 341–58.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*(4), 596–615.

Hanrieder, M. (2017). Collaborative spatial search: Implementation and validation of a multi-user task in walkable virtual environments.

Kaplan, R., Schuck, N. W., & Doeller, C. F. (2017). The role of mental maps in decision-making. *Trends in Neurosciences, 40*(5), 256–259.

Keilmann, F., de la Rosa, S., Cress, U., Schwan, S., Mohler, B. J., Buülthoff, H. H., & Meilinger, T. (2017). Collaborative cognition beats individual cognition: Advantages of collaborative over individual search require shared task information, and go along but also exceed the better distribution of labor. *Manuscript in preparation*.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science, 1*(1–12), 113–114.

Kraut, R. E., Gergle, D., & Fussell, S. R. (2002). The use of visual information in shared visual spaces: Informing the development of virtual co-presence. *In Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work* (pp. 31–40).

Löbler, H. (2010). Signs and practices: coordinating service and relationships. *Journal of Business Market Management*, *4*(4), 217–230.

Logan, R. K. (2006). The extended mind model of the origin of language and culture. *In Evolutionary Epistemology, Language and Culture*, 149–167.

Monmonier, M. (2018). *How to lie with maps*. University of Chicago Press.

Müller, R., Helmert, J. R., Pannasch, S., & Velichkovsky, B. M. (2013). Gaze transfer in remote cooperation: Is it always helpful to see what your partner is attending to? *The Quarterly Journal of Experimental Psychology*, *66*(7), 1302–1316.

Neider, M. B., Chen, X., Dickinson, C. A., Brennan, S. E., & Zelinsky, G. J. (2010). Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin, 17*(5), 718–724.

Pattee, H. H. (1997). *The physics of symbols and the evolution of semiotic controls*. Workshop on Control Mechanisms for Complex Systems: In Proc.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: an information search and protocol analysis. *Psychonomic Bulletin, 16*(2), 366–387.

Raczaszek-Leonardi, J., & Kelso, J. S. (2008). Reconciling symbolic and dynamic aspects of language: Toward a dynamic psycholinguistics. *New Ideas in Psychology, 26*(2), 193–207.

Sadalla, E. K., Burroughs, W. J., & Staplin, L. J. (1980). Reference points in spatial cognition. Reference points in spatial cognition. *Journal of Experimental Psychology: Human Learning and Memory, 6*(5), 516.

Scherbaum, S., Frisch, S., Leiberg, S., Lade, S. J., Goschke, T., & Dshemuchadse, M. (2016). Process dynamics in delay discounting decisions: An attractor dynamics approach. *Judgment and Decision Making, 11*(5), 472.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379–423.

Spranger, M. (2016). The evolution of grounded spatial language. Language Science Press.

Tomasello, M. (2014). *A natural history of human thinking*. Harvard University Press.

Zubek, J., Denkiewicz, M., Dębska, A., Radkowska, A., Komorowska-Mach, J., Litwin, P..... Raczaszek-Leonardi, J. (2016). Performance of language-coordinated collective systems: A study of wine recognition and description. *Frontiers in Psychology*, 71321.

Does the Activation of Motor Information Affect Semantic Processing?

Elisa Scerrati [ORCID]**, Cristina Iani, and Sandro Rubichi**

1 Introduction

Knowledge of object use is one of the most important available types of knowledge for a living being. For instance, humans can make use of a hammer to nail wooden planks and build a house, chimpanzees can use a twig to "fish" for insects, and birds of prey called bearded vultures, or lammergeiers, can make use of stones to break bones and feed themselves with marrow.

A basic issue in human cognition is how information concerning actions with objects is represented. Are motor representations critical components of object concepts? This question taps into the ongoing debate on the format (i.e., neural substrate, patterns of activation) of conceptual representations (for an overview see Scerrati 2017; Scerrati et al. 2017). Such debate critically involves two out of the three

E. Scerrati (✉) · S. Rubichi
Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Via A. Allegri, 9, 42121 Reggio Emilia, Italy
e-mail: elisa.scerrati@unimore.it

S. Rubichi
e-mail: sandro.rubichi@unimore.it

C. Iani
Department of Surgery, Medicine, Dentistry and Morphological Sciences with Interest in Transplant, Oncology and Regenerative Medicine, University of Modena and Reggio Emilia, Via A. Allegri, 9, 42121 Reggio Emilia, Italy
e-mail: cristina.iani@unimore.it

C. Iani · S. Rubichi
Center for Neuroscience and Neurotechnology, University of Modena and Reggio Emilia, Via G. Campi 287, 41125 Modena, Italy

main research questions outlined in the present volume, that is, how concepts become acquired and how they are being used in cognitive tasks. The current research is a psychological investigation, which attempts to address these questions and, specifically, how concept learning and representation interact with the development of motor abilities.

An increasing widespread view assumes that knowledge is grounded in sensory-motor experiences (Barsalou 1999, 2008, 2016; Gallese and Lakoff 2005; Glenberg and Kaschak 2002, 2003; Glenberg and Robertson 2000; Pulvermüller 1999, 2001; Zwaan 2004). The semantic analysis reported in Vernillo (Chap. 8) demonstrated that the literal meaning of action verbs poses constrains on their usage in metaphorical sentences. Neuropsychological research provides further support for the grounding assumption by showing the existence of selective impairments at the expenses of specific categories of information. For example, following a stroke, a viral infection or a neurodegenerative disease, such as the Alzheimer disease (AD) or Semantic Dementia (SD), people may selectively lose knowledge of living animate (i.e., animals) or inanimate (i.e., fruit/vegetables) entities, conspecifics (i.e., other people) or non-living things (i.e., manipulable artefacts). According to the *sensory/functional* theory (Warrington and McCarthy 1983, 1987; Warrington and Shallice 1984; see also Damasio 1989; Farah and McClelland 1991; Humphreys and Forde 2001; McRae and Cree's 2002), category-specific deficits can be explained by assuming that knowledge of a specific category is located near the sensory and motor areas of the brain dedicated to perception of its instances' perceptual qualities and kind of movements. Therefore, when a sensory-motor area is damaged, the processing of instances of the specific category that rely on that area is impaired. Importantly, neuropsychological research also suggests that sensory-motor representations are involved not only in comprehending and producing voluntary movements but also in thinking about them (Buxbaum et al. 2000).

In addition, neuroimaging studies have largely shown different neural activations for different categories. For instance, Chao et al. (1999, 2002) found differential activation for animals and tools. Furthermore, Chao and Martin (2000) described regions in the dorsal visual pathway, such as the posterior parietal cortex, that were differentially recruited when participants viewed manipulable objects like tools and utensils. Also, semantic knowledge of actions has been shown to involve different loci of representation in the brain than semantic knowledge of entities, specifically the frontal lobe motor-related areas (see, for example, Hickok 2014; Kemmerer 2015). Interestingly, a growing body of neuroimaging research also shows that knowledge of object use is automatically activated upon naming (Chao and Martin 2000; Chouinard and Goodale 2010), categorizing (Gerlach et al. 2002), and even passively viewing manipulable objects (Creem-Regehr et al. 2007; Grèzes et al. 2003; Vingerhoets 2008; Wadsworth and Kana 2011).

Similarly, several behavioral studies showed that semantic content influences reach-to-grasp movement responses. For instance, Gentilucci and Gangitano (1998) found that automatic word reading influenced grasping movements: Their subjects automatically associated the meaning of the word ("*corto*: *short*", "*lungo*: *long*") with the distance to cover in order to perform a grasping action and activated a motor

program for a nearer/farther object position. Glenberg and Kaschak (2002) showed that judging sensibility of sentences was easier when the movement implied by the sentence was in the same direction as the movement required by the response. In a similar vein, Zwaan et al. (2002) showed that object verification and naming was easier when the object's shape on display matched the shape implied by a previously presented sentence. Furthermore, Glover et al. (2004) demonstrated that reading words describing objects activated motor tendencies, which influenced the grasping of target blocks. Lindemann et al. (2006) further showed that action semantics activation hinges on the specific action intention of an actor. Importantly, Myung et al. (2006) showed similar effects of semantics with a lexical decision task that required keypress responses: Performance on the target word was better when semantically dissimilar prime-target pairs shared manipulation information (e.g., *typewriter* and *piano*).

Although much is known about how semantic content mediates action in response to the environment, the influence of motor activation on semantic processing did not receive as much attention. The present study aimed at filling this gap by focusing on potential effects of action on language. If, as assumed by the *sensory/functional* theory (Warrington and McCarthy 1983, 1987; Warrington and Shallice 1984; see also Damasio 1989; Farah and McClelland 1991; Humphreys and Forde 2001; McRae and Cree's 2002), conceptual content is stored closed to the sensory and motor systems, and, as claimed by the grounded view, semantics shares a common neural substrate with the sensory and the motor systems (Barsalou 1999, 2008, 2016), then effects should be observed bilaterally, that is, not only from language to action but also vice versa (see Meteyard and Vigliocco 2008).

The current study is aimed at testing whether: (a) motor information concerning objects can be pre-activated through the presentation of images of graspable objects as primes (e.g., "frying pan"); and (b) pre-activated motor information concerning graspable objects can affect performance on a lexical decision task involving target words describing objects' properties relevant for action (e.g., *handle*).

To this end, participants were instructed to observe a prime object that could be presented in two different orientations, that is, with the action-relevant component (e.g., the frying pan's handle) oriented either toward the left or toward the right. They were then asked to perform a lexical decision task (LDT)—a task commonly used in studies on lexical-semantic processing (Meyer and Schvaneveldt 1971; see also Iani et al. 2009; Scerrati et al. 2017)—on a subsequent target word. Specifically, they were required to judge whether the following target was a known word in the Italian lexicon or not by pressing a key either on the same side as the depicted action-relevant property of the prime object (i.e., spatially compatible key) or on the opposite side (i.e., spatially incompatible key). Target words matching in frequency and length were of three different types: words describing properties relevant for action with the object (action-relevant words, e.g., *handle*); words describing properties irrelevant for action with the object (action-irrelevant words, e.g., *ceramic*); words describing things unrelated to the object (unrelated words, e.g., *eyelash*).

If the image of the graspable object (i.e., the prime image) directly cues a specific motor representation, which becomes part of the concept held in working memory

(e.g., Bub and Masson 2010), then we should observe a facilitation on the subsequent lexical decision task provided that the target word is action-relevant (e.g., *handle*) and the orientation of the action-relevant component of the prime object is spatially compatible with the response key. Indeed, several behavioral studies showed a facilitation when the responding hand of the participant and the orientation of the object's graspable component, that is, its *affordance* (e.g., the handle; for the original idea of *affordance* see Gibson 1979) were compatible (i.e., on the same side) rather than incompatible (i.e., on opposite sides). This finding supports the assumption that seeing a picture of a graspable object activates the motor actions associated with its use (Iani et al. 2019; Pellicano et al. 2010; Saccone et al. 2016; Scerrati et al. 2019, 2020; Tipper et al. 2006; Tucker and Ellis 1998; Vainio et al. 2007). Therefore, we expect that the presentation of the graspable prime object will pre-activate manipulation information about objects. This in turn should facilitate a lexical decision task on target words describing those objects' properties relevant for action (e.g., *handle*). In contrast, no such facilitation is expected for target words that describe properties irrelevant for action with (action-irrelevant words, e.g., *ceramic*) or unrelated to (unrelated words, e.g., *eyelash*) the prime object. In other words, we expect that motor information evoked by object observation will have different effects as a function of the following type of word. Specifically, we predict that motor information will determine a motor-to-semantic priming effect for action-relevant words as the processing of these words can benefit from the activation of motor knowledge. Conversely, it should determine neither benefits nor disadvantages for action-irrelevant and unrelated words as these words refer to motor-irrelevant features of the prime objects. Hence, we expect to observe an interaction between spatial compatibility and the type of word.

2 Method

2.1 Materials

The prime stimuli were digital photographs of four domestic objects (can, door, frying pan, radiator) selected from public-domain images available on the Internet. Prime objects could be presented in two orientations, that is, with the action-relevant component (e.g., the frying pan's handle) oriented either toward the left or toward the right. These objects subtended a maximum of 13.7° of visual angle horizontally and 12.3° of visual angle vertically when viewed from a distance of 60 cm. Prime objects were centered on screen according to the length and width of the entire object.

The target stimuli were twelve words belonging to three different categories: Four words referred to a characteristic of the prime object that was relevant for action (e.g., *handle*); four words referred to a characteristic of the prime object that was irrelevant for action (e.g., *ceramic*); four words referred to things unrelated to the prime object (e.g., *eyelash*). For the complete list of stimuli, see Appendix. Target words ranged

Table 1 Psycholinguistic matched variables of the target words used in the main experiment

	Action-relevant words			Action-irrelevant words			Unrelated words		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Frequency	7.7	3.7	3–12	21.2	11.5	11–34	26	40.7	2–87
Length	7.7	1.2	6–9	8	2.1	5–10	7.2	0.9	6–8

from 2.7 to 5.4 cm (from 5 to 10 characters) which resulted in a visual angle range between 2.5° and 5.1° when viewed from a distance of 60 cm.

Words from the three categories (action-relevant, action-irrelevant, and unrelated) were matched in terms of frequency and length. For lexical frequency, the Italian database Colfis was used (Bertinetto et al. 1995). Values for frequency and length of target words are reported in Table 1.

To control for association strength between the prime object and the target word, 40 Italian participants (23 males; mean age: 28 years old; SD: 9 years) who did not participate in the main Experiment were asked to rate the twelve target words in terms of their degree of association with the prime objects on a 1–7 points Likert scale (1 = "not associated at all"; 7 = "very associated"). The mean ratings were 5.2 for action-relevant words related to the prime object, 5.4 for action-irrelevant words related to the prime object, and 1.5 for words unrelated to the prime object.

Twelve legal non-word fillers (e.g., *celimora*) were created using a non-word generator for the Italian language available online.[1] The non-words were preceded by the same prime objects.

To control for potential phonological associations between the non-word fillers and the target words, 28 new Italian participants (11 males; mean age: 27 years old; SD: 7 years) were engaged in a free association production task. The task required participants to write down the first two Italian words that each of the twelve non-words brought to mind. Only one participant reported the Italian word *ciglia* (included in the unrelated category) in response to the non-word *geglie*. However, given it was an isolated case, we did not consider it necessary to exclude this non-word from our selection of non-word fillers.

3   Participants

Thirty-four participants (13 males; mean age: 22 years old; SD: 3 years) from the University of Modena and Reggio Emilia where the experiment was conducted. All participants were native speakers of Italian, had normal or corrected to normal vision, and were naïve as to the purpose of the experiment. Handedness was measured by the Edinburgh Handedness Inventory (Oldfield 1971), which revealed that 25 participants were right-handed (laterality mean = 0.76; SD = 0.13), seven participants were

---

[1]https://www.trainingcognitivo.it/GC/nonparole/.

ambidextrous (laterality mean $= 0.25$; SD $= 0.21$) and two participants were left-handed (laterality mean $= -0.69$; SD $= 0.10$). The experiment was conducted in accordance with the ethical standards laid down in the Declaration of Helsinki and fulfilled the ethical standard procedure recommended by the Italian Association of Psychology (AIP). All procedures were approved by the Department of Education and Human Sciences of the University of Modena and Reggio Emilia where the experiment was conducted. All participants gave their written informed consent to participate to the study.

4 Apparatus

Stimulus presentation, response times (RTs) and accuracy were controlled and recorded by E-Prime 2 (Psychology Software Tools, Inc., Sharpsburg, PA). Participants completed the experiment on a HP ProDesk 490 G1 MT running Windows 7 with a 19 *in* monitor and a display with a resolution of 1280×1024 pixels.

5 Design and Procedure

Two factors were manipulated: *Target word* with 3 levels (action-relevant; action-irrelevant; unrelated), and *Spatial compatibility*—between the orientation of the action-relevant component of the prime object and the response—with two levels (spatially compatible: both handle and response on the right or on the left; spatially incompatible: handle on the right and response on the left and *viceversa*). Both factors were manipulated within-subject.

Participants sat at a viewing distance of about 60 cm from the monitor in a dimly-lit room. Each trial started with the presentation of a fixation cross (0.3 cm $\times 0.3$ cm) for 500 ms. Immediately after the fixation, the prime object appeared on screen for 1000 ms. Then, either the target word or the non-word filler was displayed on screen until a response was given or until 1500 ms had elapsed (see Fig. 1 for details). RT latencies were measured from the onset of the target stimulus. Both target and filler stimuli were bold lowercase Courier new 18 and were presented in black in the center of a white background.

Participants were asked to make a lexical decision, that is, determine whether the displayed letter string was an Italian word or not, by pressing one of two lateralized buttons as quickly and as accurately as possible. Response keys were the "-" and the "z" keys on an Italian QWERTY keyboard. Half of the participants responded by pressing the "-" key with their right index finger when the letter string was an Italian word, and the "z" key with their left index finger when it was a non-word. The other half was assigned to the opposite mapping.

The order of presentation of each prime-target pair was randomized across participants. The experiment consisted of 24 practice trials (different from those used in
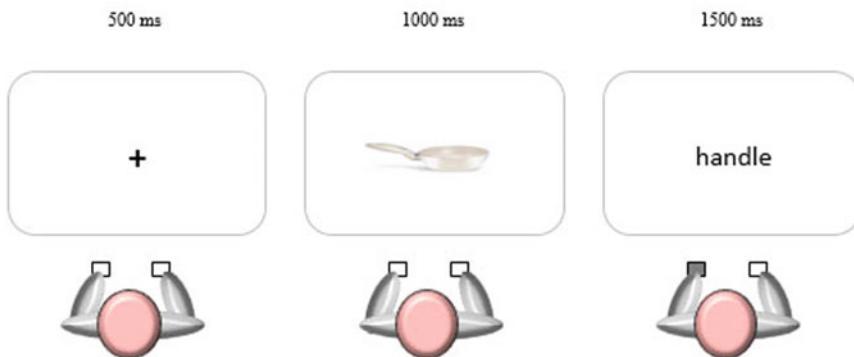
Fig. 1 Illustration of an action-relevant target word in the spatially compatible condition. In the example above instructions required to respond with the left index finger to words and with the right index finger to non-words. Note that elements are not drawn to scale

the experiment) and two experimental blocks of 48 trials each, for a total of 120 trials per participant. Blocks were separated by a self-paced interval and the experiment lasted approximately 10 min.

6 Results

Responses to non-word fillers were discarded. Omissions (1%) and outlying RT (5%) that were two standard deviations (SD) from the participant's mean were excluded from the analysis.

Two repeated measures ANOVAs with *Target Word* (action-relevant, action-irrelevant, unrelated) and *Spatial compatibility* (compatible, incompatible) as within-subject factors were conducted, one for RT latencies and one for percentage errors (3.5%). When sphericity was violated, the Huynh–Feldt correction was applied, although the original degrees of freedom are reported.

The results of the ANOVA on the RT latencies did not reveal any significant main effect or interaction, all $F < 1$. In contrast, the results of the ANOVA on the percentage errors showed a significant main effect of *Target Word* ($F(2, 66) = 3.67$, $MS_e = 61.15$, $p = 0.043$, $n_p^2 = 0.10$), that is, lexical decision responses were more accurate for action-relevant target words (1.65%) than for both action-irrelevant (4.22%) and unrelated target words (4.59%), $t(33) = 2.92$, $p = 0.006$, and $t(33) = 2.61$, $p = 0.01$, respectively. No other main effect resulted significant, $F < 1$. Results are shown in Fig. 2.

Importantly, there was a marginally significant interaction between *Target Word* and *Spatial compatibility* ($F(2, 66) = 3.42$, $MS_e = 35.68$, $p = 0.057$, $n_p^2 = 0.09$). Paired comparisons revealed that lexical decision responses for action-relevant target words tended to be more accurate in the spatially compatible condition (0.73%) than
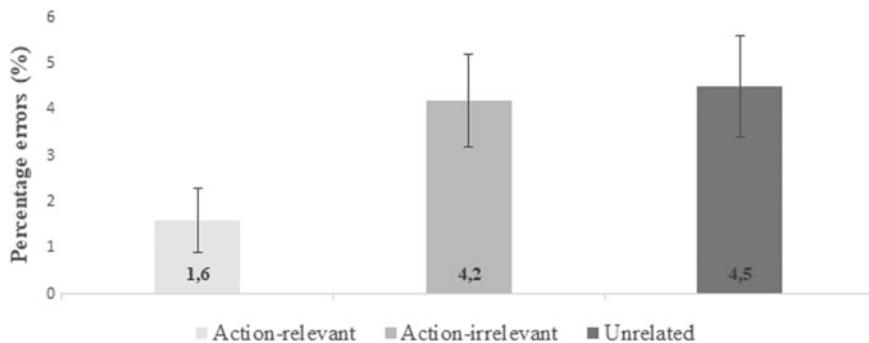
Fig. 2 Mean lexical decision percentage errors as a function of target word (action-relevant; action-irrelevant; unrelated): bars indicate standard errors
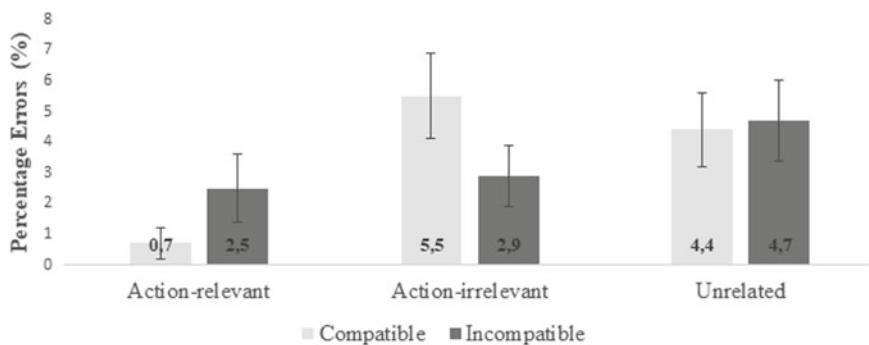
Fig. 3 Mean lexical decision percentage errors as a function of target word (action-relevant; action-irrelevant; unrelated) and spatial compatibility (compatible; incompatible): bars indicate standard errors

in the spatially incompatible condition (2.57%), $t(33) = 1.71$, $p = 0.09$ two tailed. In contrast, lexical decision responses for action-irrelevant target words tended to be more accurate in the spatially incompatible condition (2.94%) than in the spatially compatible condition (5.51%), $t(33) = -1.74$, $p = 0.09$ two tailed. Finally, lexical decision responses for unrelated target words did not differ in the spatially compatible (4.41%) and incompatible (4.77%) conditions. Figure 3 shows the results graphically.

7 Discussion

Although much evidence is available on the influence of semantics on action preparation and execution (Gentilucci and Cangitano 1998; Glenberg and Kaschak 2002; Glover et al. 2004; Lindemann et al. 2006; Myung et al. 2006; Zwaan et al. 2002), the effects of motor control on language processing are poorly investigated.

The current study examined whether semantic processing may be influenced by the activation of the motor system. If conceptual content is stored closed to the sensory and motor systems (Warrington and McCarthy 1983, 1987; Warrington and Shallice 1984; see also Damasio 1989; Farah and McClelland 1991; Humphreys and Forde 2001; McRae and Cree's 2002), and if it shares a common neural substrate with the sensory and the motor systems (Barsalou 1999, 2008, 2016), then effects of language on action and of action on language should be observed likewise (Meteyard and Vigliocco 2008).

We explored whether presenting images of graspable objects (e.g., "frying pan") as prime stimuli could pre-activate manipulation information about objects, which in turn could facilitate a lexical decision task on target words referring to objects' properties relevant for action (e.g., *handle*). That is, we expected that object observation would activate motor knowledge leading to a motor-to-semantic priming effect only for target words referring to action-relevant components of objects as only the processing of action-relevant words should benefit from the activation of motor knowledge.

In line with our hypothesis, we found that performing a lexical decision on action-relevant target words produced more accurate responses than performing the same task on action-irrelevant words and on words unrelated to the prime objects. This finding suggests that language processing is somewhat facilitated provided that words are not only related to the prime object seen before but also relevant for action with that object. It is plausible to assume that the prime object's graspability was able to shift participants' attention to the action-relevant features of the object thus facilitating the subsequent lexical decision on words describing those features.

Furthermore, we found an interaction between the type of word (relevant-for-action; irrelevant-for-action; unrelated) and spatial compatibility (compatible, incompatible). In line with our hypothesis, we observed a tendency toward lower percentage errors (i.e., facilitation) when the target word was action-relevant (e.g., *handle*) and there was spatial compatibility between the orientation of the action-relevant component of the prime object and the response. Conversely, we observed a tendency toward higher percentage errors (i.e., interference) when the target word was action-irrelevant (e.g., *ceramic*) and there was spatial compatibility between the orientation of the action-relevant component of the prime object and the response. Therefore, motor information activated by observing objects' orientation may influence language processing to the extent that words being processed are relevant for action with such objects. This preliminary finding supports the assumption that observing a graspable object activates the motor actions associated with its use (Iani et al. 2019; Pellicano et al. 2010; Saccone et al. 2016; Scerrati et al. 2019, 2020; Tipper et al. 2006; Tucker and Ellis 1998; Vainio et al. 2007).

Taken together these findings suggest that the activation of motor information may affect semantic processing.

However, the present study has a limitation in that our results only emerged for percentage errors (not response latencies). This may be the consequence of the low level of verbal processing involved by the lexical decision task. Indeed, the LDT may recruit the semantic system to a small extent (see Scerrati et al. 2017) thus

failing to show a robust influence of motor information on language processing that is able to affect response latencies (for task-dependent influences of motor information on conceptual processing see De Bellis et al. 2016; see García and Ibáñez 2016 for review). That is, if the LDT is performed by relying on a simple word association strategy, i.e., without determining the type of association between the property word and the concept word (for example, whether the property word refers to a part of the concept word as in the concept-property pair *frying pan-handle*), then the underlying conceptual representations may not be retrieved at all, this resulting in motor information being unable to exert a robust influence on semantic processing (e.g., Solomon and Barsalou 2004). In addition, as highlighted by a recent review by García and Ibáñez (2016), the allowed time-lag (2.5 s) between motor and linguistic information may have played a role in our study leading to a weaker influence of motor knowledge on language processing. Such weakened influence may reflect in the motor-to-semantic priming effect failing to show for response latencies. Even holding these caveats in mind, our study indicates a possible influence of motor control on cognitive functions and strengthens the hypothesis of the proximity of language and sensory-motor systems in the human brain (see also Goldstone and Barsalou 1998).

Future studies may extend the investigation of mutual effects of semantic content and motor control by introducing other tasks that more explicitly require the construction of modality-specific representations (e.g., motor representations). In fact, it is plausible that a conceptual, recognition-oriented task may reveal effects of motor control on semantic processing more easily than a more implicit task such as the lexical decision task. A different task will help identify to which extent the nature of the task determines the motor-to-semantic priming effect and to discard other possible factors.

Appendix

Prime objects	Target words			Non-words		
	Action-relevant	Action-irrelevant	Unrelated			
calorifero *radiator*	manopola *knob*	ghisa *cast iron*	panchina *bench*	agraccia	bucconede	celimora
lattina *can*	linguetta *tab*	alluminio *aluminium*	astuccio *case*	conichia	fangialle	geglie
padella *frying pan*	manico *handle*	ceramica *ceramic*	corredo *dowry*	ghipi	naseco	mezecolo

(continued)

Prime objects	Target words			Non-words		
	Action-relevant	Action-irrelevant	Unrelated			
porta *door*	maniglia *door-handle*	compensato *plywood*	ciglia *eyelash*	ommibicio	rinchite	sobbeme

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–660.

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

Barsalou, L. W. (2016). On staying grounded and avoiding quixotic dead ends. *Psychonomic Bulletin & Review, 23*, 1122–1142.

Bertinetto, P. M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., & Thornton, A. M. (1995). CoLFIS (Corpus e Lessico di Frequenza dell'Italiano Scritto) [Corpus and frequency lexicon of written Italian]. *Institute of Cognitive Sciences and Technologies.*

Bub, D. N., & Masson, M. E. (2010). On the nature of hand-action representations evoked during written sentence comprehension. *Cognition, 116*(3), 394–408.

Buxbaum, L. J., Veramontil, T., & Schwartz, M. F. (2000). Function and manipulation tool knowledge in apraxia: knowing 'what for' but not 'how'. *Neurocase, 6*(2), 83–97.

Chao, L. L., & Martin, A. (2000). Representation of manipulable manmade objects in the dorsal stream. *NeuroImage, 12*, 478–484.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in posterior temporal cortex for perceiving and knowing about objects. *Nature Neuroscience, 2*, 913–919.

Chao, L. L., Weisberg, J., & Martin, A. (2002). Experiencedependent modulation of category related cortical activity. *Cerebral Cortex, 12*, 545–551.

Chouinard, P. A., & Goodale, M. A. (2010). Category-specific neural processing for naming pictures of animals and naming pictures of tools: An ALE meta-analysis. *Neuropsychologia, 48*(2), 409–418.

Creem-Regehr, S. H., Dilda, V., Vicchrilli, A. E., Federer, F., & Lee, J. N. (2007). The influence of complex action knowledge on representations of novel graspable objects: Evidence from functional magnetic resonance imaging. *Journal of the International Neuropsychological Society, 13*(6), 1009–1020.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition, 33*, 25–62.

De Bellis, F., Ferrara, A., Errico, D., Panico, F., Sagliano, L., Conson, M., & Trojano, L. (2016). Observing functional actions affects semantic processing of tools: Evidence of a motor-to-semantic priming. *Experimental Brain Research, 234*(1), 1–11.

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General, 120*(4), 339–357.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory–motor system in conceptual knowledge. *Cognitive Neuropsychology, 22*(3/4), 455–479.

García, A. M., & Ibáñez, A. (2016). A touch with words: Dynamic synergies between manual actions and language. *Neuroscience & Biobehavioral Reviews, 68*, 59–95.

Gentilucci, M., & Gangitano, M. (1998). Influence of automatic word reading on motor control. *European Journal of Neuroscience, 10*(2), 752–756.

Gerlach, C., Law, I., & Paulson, O. B. (2002). When action turns into words. Activation of motor-based knowledge during categorization of manipulable objects. *Journal of Cognitive Neuroscience, 14*(8), 1230–1239.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston, MA: Houghton Mifflin.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*(3), 558–565.

Glenberg, A. M., & Kaschak, M. P. (2003). The body's contribution to language. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 93–126). San Diego, CA: Academic Press.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language, 43*, 379–401.

Glover, S., Rosenbaum, D. A., Graham, J., & Dixon, P. (2004). Grasping the meaning of words. *Experimental Brain Research, 154*(1), 103–108.

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition, 65*(2–3), 231–262.

Grèzes, J., Tucker, M., Armony, J., Ellis, R., & Passingham, R. E. (2003). Objects automatically potentiate action: An fMRI study of implicit processing. *European Journal of Neuroscience, 17*(12), 2735–2740.

Hickok, G. (2014). *The myth of mirror neurons: The real neuroscience of communication and cognition.* New York, NY: Norton.

Humphreys, G. W., & Forde, E. M. E. (2001). Hierarchies, similarity, and interactivity in object recognition: "Category-specific" Neuropsychological deficits. *Behavioral and Brain Sciences, 24*(3), 453–509.

Iani, C., Ferraro, L., Maiorana, N. V., Gallese, V., & Rubichi, S. (2018). Do already grasped objects activate motor affordances? *Psychological Research, 83*, 1363–1374, https://doi.org/10.1007/s00426-018-1004-9.

Iani, C., Job, R., Padovani, R., & Nicoletti, R. (2009). Stroop effects on redemption and semantic effects on confession: Simultaneous automatic activation of embedded and carrier words. *Cognitive Processing, 10*, 327–334.

Kemmerer, D. (2015). Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. *Psychonomic Bulletin & Review, 22*, 1068–1075.

Lindemann, O., Stenneken, P., Van Schie, H. T., & Bekkering, H. (2006). Semantic activation in action planning. *Journal of Experimental Psychology: Human Perception and Performance, 32*(3), 633.

McRae, K., & Cree, G. S. (2002). Factors underlying category-specific semantic deficits. In E. M. E. Forde & G. Humphreys (Eds.), *Category-specificity in brain and mind* (pp. 211–249). East Sussex, UK: Psychology Press.

Meteyard, L., & Vigliocco, G. (2008). The role of sensory and motor information in semantic representation: A review. In P. Calvo, & T. Gomila (Eds.), *Handbook of cognitive science: An embodied approach* (pp. 291–312). Elsevier.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227.

Myung, J. Y., Blumstein, S. E., & Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: Manipulation knowledge of objects. *Cognition, 98*(3), 223–243.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113.

Pellicano, A., Iani, C., Borghi, A. M., Rubichi, S., & Nicoletti, R. (2010). Simon-like and functional affordance effects with tools: The effects of object perceptual discrimination and object action state. *Quarterly Journal of Experimental Psychology, 63*(11), 2190–2201.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences, 22*, 253–336.

Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences, 5*(12), 517–524.

Saccone, E. J., Churches, O., & Nicholls, M. E. (2016). Explicit spatial compatibility is not critical to the object handle effect. *Journal of Experimental Psychology: Human Perception and Performance, 42*(10), 1643–1653.

Scerrati, E. (2017). *From amodal to grounded to hybrid accounts of knowledge: New evidence from the investigation of the modality-switch effect*. Unpublished doctoral dissertation, University of Bologna, Bologna, Italy.

Scerrati, E., Iani, C., Lugli, L., & Rubichi, S. (2019). C'è un effetto di potenziamento dell'azione con oggetti bimanuali? *Giornale Italiano di Psicologia, 4*, 987–996, https://doi.org/10.1421/95573.

Scerrati, E., Iani, C., Lugli, L., Nicoletti, R., & Rubichi, S. (2020). Do my hands prime your hands? The hand-to-response correspondence effect? *Acta Psychologica, 203*, 103012, https://doi.org/10.1016/j.actpsy.2020.103012.

Scerrati, E., Lugli, L., Nicoletti, R., & Borghi, A. M. (2017). The multilevel modality-switch effect: what happens when we see the bees buzzing and hear the diamonds glistening. *Psychonomic Bulletin & Review, 24*(3), 798–803.

Solomon, K. O., & Barsalou, L. W. (2004). Perceptual simulation in property verification. *Memory & Cognition, 32*(2), 244–259.

Tipper, S. P., Paul, M. A., & Hayes, A. E. (2006). Vision-for-action: The effects of object property discrimination and action state on affordance compatibility effects. *Psychonomic Bulletin & Review, 13*(3), 493–498.

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 830–846.

Vainio, L., Ellis, R., & Tucker, M. (2007). The role of visual attention in action priming. *The Quarterly Journal of Experimental Psychology, 60*(2), 241–261.

Vingerhoets, G. (2008). Knowing about tools: Neural correlates of tool familiarity and experience. *Neuroimage, 40*(3), 1380–1391.

Wadsworth, H. M., & Kana, R. K. (2011). Brain mechanisms of perceiving tools and imagining tool use acts: A functional MRI study. *Neuropsychologia, 49*(7), 1863–1869.

Warrington, E. K., & McCarthy, R. A. (1983). Category specific access dysphasia. *Brain, 106*, 859–878.

Warrington, E. K., & McCarthy, R. A. (1987). Categories of knowledge: Further fractionations and an attempted integration. *Brain, 110*, 1273–1296.

Warrington, E. K., & Shallice, T. (1984). Category-specific semantic impairments. *Brain, 107*, 829–854.

Zwaan, R. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation 44* (pp. 35–62). San Diego, CA: Academic Press.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168–171.

Grounding Abstract Concepts in Action

Semantic Analysis of Four Italian Action Verbs Encoding Force Events

Paola Vernillo

1 Introduction

Over the past recent years, different works connected over the idea that language, cognition, and bodily experience must be considered as inextricably intertwined areas of research (Gallese and Lakoff 2005; Lakoff and Johnson 1999). A consistent number of multidisciplinary studies showed that sensory-motor information influences our cognitive structures and thus represents a primary source in the operation of meaning construction (amongst others, Martin and Chao 2001; Pulvermüller 2005).

In this large frame, action verbs play a pivotal role. They are recognized as primary tools both in the linguistic encoding of bodily knowledge and in the linguistic representation and modeling of a wide array of highly abstract concepts (Panunzi and Vernillo 2019). Action verbs are mainly used to encode very different types of action events and bodily schemas. Their semantic extension allows us to refer to a myriad of experiences, affordances, bodily movements, and relations between physical objects (i.e., primary variation). Moreover, these predicates are pervasively used to encode a large and complex array of abstract concepts and figurative meanings (i.e., marked variation), for whose labeling they coherently re-use their rich action imagery. The class of action verbs represents a case of exceptional interest within the verb lexicon category. These verbs are not only among the primary words of children's vocabulary (Tomasello 2003) but they are also among the most common tools in oral communication, having an even more significant weight than nouns in spoken language use

P. Vernillo (✉)
Università degli Studi di Firenze (UNIFI), Firenze, Italy
e-mail: paola.vernillo@unifi.it

167

(Gagliardi 2014; Moneglia 2014a, b). Moreover, and differently from other predicates, action verbs directly anchor, on the level of language, the domain of sensory-motor experience to that of highly abstract thought. Therefore, the analysis of these predicates' semantic variation may ease the understanding of how spatial and bodily information (spatial vectors, motion patterns, force dynamics) is mapped to make new and non-literal meanings emerge.

This work,[1] whose primary research field is that represented by Cognitive Linguistics and Semantic Theory, starts from the hypothesis that there exists a sort of hidden relation between the two dimensions of use and meaning of a given action verb (i.e., *primary* and *marked variation*); and that it there also exists a sort of correspondence between the type of action and metaphorical concepts which can be expressed by means of the same predicate. The main research questions this study has been built upon can be spelled out as it follows:

1. What are the relationships between the concrete (i.e., primary variation) and the metaphorical (i.e., marked variation) uses of a given action verb?
2. Which semantic features of the action verb determine the metaphorical potential of the verb? And how do we determine which action verb can allow us to access which metaphorical concept or figurative meaning?
3. Finally, how can we explain divergent metaphorical potentials of action verbs involved in the encoding of the same type of action events (i.e., locally equivalent verbs)?

It is worth to bear in mind that these questions are not only relevant with respect to my research field (i.e., Linguistics), but are closely connected to the three main research questions this volume starts from (Bechberger and Liu, this volume):

a. On the representation level: how can we formally describe and model concepts (Färber, Svetashova, and Harth, this volume; Gust and Umbach, this volume)? And, more specifically, how do we use characteristics of action concepts to formally model more abstract ones?
b. On the learning level: where do concepts come from and how are they acquired (Bechberger and Kühnberger, this volume)? And, in particular, how do we transfer sensory-motor information to new and more abstract contexts?
c. On the application level: how are concepts used in cognitive tasks (Gega, Liu and Bechberger, this volume; Scerrati, Iani, and Rubichi, this volume; Schneider and Nürnberger, this volume)? And, with respect to the present research, how do we apply our action-knowledge in linguistic contexts where no physical action is implied?

To give all these questions an answer, in this study, I aim at investigating the semantic variation of a small group of Italian action verbs (ita., *premere*, *spingere*, *tirare* and *trascinare*; Eng., *to press*, *to push*, *to pull*, and *to drag*) involved in the encoding of the force-dynamics category (Langacker 1987). Although the four verbs

---

[1]This research partially bases on two previous works (Panunzi and Vernillo 2019; Vernillo 2019) and on the author's doctoral dissertation (Vernillo 2020, unpublished).

in analysis belong to the same semantic class (i.e., force), they profile different types of action concepts and events. It seems reasonable to believe that the specific image-schematic features associated with their action imagery influence the differences in their semantic extension and linguistic use. Nevertheless, along the semantic axis (i.e., primary and marked variation), there also exist specific points where the uses of these verbs tend to converge. For instance, happens that, in some specific pragmatic contexts, the uses of *premere* converge with those of *spingere* (e.g., setting relationships between objects), or the uses of *trascinare* converge with those of *tirare* (e.g., the frictional motion of an object along a surface). These verbs show a partial convergence (or divergence) not only with respect to their primary variation (i.e., when encoding physical concrete meanings), but also with respect to their marked variation (i.e., when encoding figurative meanings). For example, there are cases in which the verb *premere* (Eng., *to press*) and the verb *spingere* (Eng., *to push*) refer to the same type of metaphorical concept (e.g., PSYCHOLOGICAL FORCES ARE PHYSICAL FORCES), or cases in which the verb *tirare* (Eng., *to pull*) and *trascinare* (Eng., *to drag*) encode the same type of conceptual metaphor (e.g., CAUSES ARE FORCES AFFECTING MOTION).

The present study bases on the idea that a deep analysis of the action imagery associated to these predicates can help us to shed new light on their behavior in metaphorical contexts. To support this idea, in the following paragraphs, I will describe the semantic variation of each of the four verbs, mainly focusing on the salient image-schematic structures and the specific action schemas that characterize the primary core of the verbs. Additionally, I will explain how these same structures and schemas permit to bond together the *marked* (i.e., largely metaphorical) and the *primary variation* of the verbs (Lakoff 1990, 1993; Turner 1991). In Sect. 2, I will present the ontological infrastructure within which my analysis was developed with. In Sect. 3, I will give a general overview of the theoretical approaches (i.e., Conceptual Metaphor Theory, Image Schema Theory, and Embodiment) that mainly influenced my approach to the analysis of action verbs. In Sect. 4, I will present the collection of data and the methodology I used for these predicates' annotation. Section 5 will describe the primary variation of each of the four verbs, and it will be mainly focused on the salient image-schematic structure and the specific action schemas that characterize the primary core of the verbs. In Sect. 6, I will illustrate the marked variation of the predicates, and I will explain how the same structures and schemas highlighted in the primary variation of the four verbs permit the bonding of the marked (or largely metaphorical) and the primary variation of the verbs (see the Invariance principle: Lakoff 1990, 1993; Turner 1991). In Sect. 7, I will briefly discuss the results obtained by comparing primary and marked variation of the four predicates in the analysis. First, I will show that the results of the study are consistent with the idea that metaphorical extensions of action verbs are constrained by the image-schematic structures involved in the core meaning of the verbs. Second, I will point out that these same structures are also responsible for the divergencies found within the metaphorical variation of action verbs pertaining to the same semantic class (i.e., force). Finally, in Sect. 8, I will draw some general conclusions about the type of study that I proposed.

Table 1 Visual representation of action concepts in IMAGACT

Action type	Scene	Standardization	Verb	LEV
Setting relations between two objects		Marta preme il coperchio sulla scatola (Eng., Mary presses the lid onto the box)	*Premere* (Eng., *to press*)	*Spingere* (Eng., *to push*)

2 The Semantic Representation of Action Verbs in the IMAGACT Ontology

The semantic characterization of action verbs given in this study owes a great deal to the representation of action events and concepts the IMAGACT Ontology was built upon. This is why, the following paragraphs will be devoted to the general description of the Ontology (Sect. 2.1), and the definition of the notion of primary (Sect. 2.2) and marked variation (Sect. 2.3).

2.1 The Internal Structure of the IMAGACT Ontology

IMAGACT is a multimodal and multilingual ontology that depicts action via a visual representation system. The choice to represent action concepts by using both prototypical 3D animations and brief videos (Moneglia 2014a, b; Panunzi et al. 2014) stemmed from two needs: first, to avoid the vagueness of semantic definitions, and second, to have a resource that could have disentangled action categorization from a specific language representation (Brown 2014).

IMAGACT includes more than 1000 distinct action scenes that have been primarily derived from the annotation of spoken language corpora in English and Italian. While in a preliminary phase of the project, Chinese and Spanish data were also processed, extensions to (Syrian) Arabic, Danish, German, Hindi, Japanese, Polish, Portuguese, and Serbian were made available on the online interface[2] only recently.

The visual representation of the action concepts is organized as the following: each prototypical scene is linked to a single action concept (or action type), each action verb is connected to more than one prototypical scene, and each prototypical scene is associated to more than one action verb. Some action verbs share a common referent (or a subset of action scenes) and are hence called locally equivalent verbs (e.g., *to push* and *to press*). The following Table 1 gives a brief schematization.

Concerning the present analysis, the IMAGACT framework represents an important point of reference for the investigation of action verbs semantics. First, the

---

[2]https://www.imagact.it/imagact/query/dictionary.seam.

Ontology contains a consistent amount of data that has been massively taken from multiple spoken resources (e.g., IMAGACT and BNC corpora). Second, this resource provides a well-structured visual categorization of action concepts and of bodily schemas encoded by general action verbs which are most used in everyday language. Third, it permits to hook the linguistic representation of the highly abstract concepts (and of the figurative meanings) encoded by a given action verb to the very inherent semantic core of the verb. Finally, it eases the interpretation of the variation axes of the predicate (i.e., primary and marked variation), since they are jointly considered rather than entirely separate dimensions of the lexical item. The rich semantic information included in the database helped to better structure the annotation of the metaphorical uses of the action predicates. Moreover, it helped to expand the number of details that have been used to show that either metaphorical and physical uses of action verbs are not randomly produced, but that they both refer to crucial motor and perceptual inputs coming from our cognitive and actual representation of actions.

2.2   The Primary Variation of Action Verbs

Within the IMAGACT Ontology, the semantics of action is described as based two main axes of variation: the primary and the marked variation. Importantly, the resource keeps the verb occurrences of the two types of variation well distinct. The procedure via which metaphorical and phraseological usages are separated from those strictly referring to physical actions is made possible through the adoption of an operational *test à la Wittgenstein* (Gagliardi 2014). According to this test, the verb uses are judged primary if it is possible to point to a certain (perceptible) event and says to someone who does not know the meaning of a given verb that "this action and similar events are what we refer to with this verb"; contrarily, the occurrences that do not instantiate the basic meaning of the verb are tagged as marked.

Within the ontology, the expression *primary variation* refers to the set of different action types to which a given action verb can refer in its proper sense (or concrete physical meaning). To illustrate this point, some of the possible physical uses of the Italian action verb *spingere* (e.g., *to push*) are considered:

(1)   "Marta spinge il pulsante"
              "Marta pushes the button"
(2)   "Marta spinge il coperchio sulla scatola"
              "Marta pushes the lid onto the box"
(3)   "Marta spinge il carrello lungo il corridoio"
              "Marta pushes the cart down the hall"
(4)   "La nuotatrice si spinge con le gambe"
              "The swimmer pushes herself off of the wall with her legs"

All the listed examples (1–4) are recognized as instantiations of the primary meaning of the verb *spingere* (Eng., *to push*). The semantics of the predicate is shown in all

its complexity while encoding different linguistic and cognitive traits. In examples (1–2), the verb can be substituted by the same locally equivalent verb (e.g., *premere),* even though the scenes refer to two action types: in the former case, the verb describes the application of force on an object to activate a connected device; in the latter case, the verb describes a situation in which a human agent applies a force to set a relation between two entities. In examples (3–4), the verb *spingere* cannot be substituted by the same locally equivalent verb: the meaning of case (3) cannot be encoded by another predicate. The event in case (4) cannot be named by a single verb but only by a more complex syntactic structure, such as 'darsi una spinta' (Eng., 'to give yourself a push'). Moreover, the examples in (3–4) describe two types of motion event in the physical space. In (3), the verb names an event in which a human agent causes an object to move along a path (caused motion). In example (4), the verb encodes an event in which a human entity moves spontaneously along a path without the intervention of an external force (self-propelled motion).

2.3   The Representation of the Marked Variation of Action Verbs

The term *marked variation* refers to the set of uses in which the action verb does not encode physical concepts but abstract/figurative (Moneglia et al. 2012; Panunzi and Moneglia 2004). Let us consider the following four sentences, which partially exemplify the variation of the verb *spingere* (e.g., *to push*):

(5)   "L'oratore spinge sui temi sociali"
              "The speaker is pressing on the social agenda"
(6)   "La situazione spinge il Consiglio a intervenire"
              "The situation is pushing the Council to intervene"
(7)   "L'autore ama spingere i suoi personaggi"
              "The author likes to push his characters"
(8)   "La situazione si spinge verso l'anno successivo"
              "The situation will extend onto the next year"

The sentences in (5–8) do not instantiate the basic meaning of the verb *spingere* (Eng., *to push*). These examples are based on different semantic processes (mostly metaphorical). Thereby the verb undergoes a semantic shift; it has thus been used to express different kinds of metaphorical meanings. In particular, the predicate represents a situation in which a speaker conveys a specific communicative intention (5), implies an act of psychological influence (6), defines the artistic manipulation put in place by an author (7), or names the time extension in the duration of an event (8).

    As stated above, marked uses are sharply separated from the occurrences referring to concrete physical actions and annotated in a different online interface. Unfortunately, although an ad-hoc infrastructure was designed to classify the marked uses

found in the variation of the action verbs (Brown 2014), the IMAGACT ontology only specifies the semantic interpretations of predicates with respect to their physical actions: hence, other kinds of interpretations are ignored and are not visually represented. The lack of a clear depiction of marked uses is not connected in any way to their semantic load within the infrastructure (they represent half of the IMAGACT database occurrences). This problem must be rather explained by reference to the visual format of the ontology, which makes it not easy to represent abstract concepts (Brown 2014).

3 Body, Metaphors, and Metaphorical Projections of Image Schemas

The analysis focuses on two essential aspects: first, the action verbs semantics and, second, the particular role played by action bodily information. In the following paragraphs, I will give a brief overview of the main theoretical scenarios my analysis has been developed with. Before going through a proper analysis of the semantic variation of action verbs, three fundamental frameworks need to be illustrated: in Sect. 3.1, I will present the main tenets underpinning the embodied theory of language. In Sect. 3.2, I will introduce the key points behind Lakoff and Johnson's Conceptual Metaphor Theory. In the final Sect. 3.3, I will focus on the Image schemas Theory, as well as its role within a deep level language analysis.

3.1 The Embodied Paradigm

Recently, interest has grown in the idea that language and cognition should be investigated with respect to the deep relationship to bodily experience (Aziz-Zadeh and Damasio 2008; Desai et al. 2011; Gallese and Lakoff 2005; Kiefer and Pulvermüller 2011; Martin and Chao 2001; Pulvermüller 2005). Embodied cognition theories are based on the assumption that between the level of cognitive processes (action and perception) and abilities (abstract thought and language comprehension) there is no defined boundary or sharp separation (Zipoli Caiani 2011). Accepting that not only the brain but also features of the agent's body play a significant role in cognitive processing means to embrace the idea that our entire conceptual system is largely constrained by the kind of body and sensory-motor processes we are characterized by as humans. The body emerges as a crucial locus and represents a functional restraint that imposes its structure on different domains of human experience (Zipoli Caiani 2011). But what does it mean to embrace the embodied paradigm when it comes to language? The embodied approaches emerged in response to the cartesian (or cognitivist) paradigm. According to this paradigm, the brain is viewed as a processor

of abstract information, while cognition should be defined as the computation of abstract symbols that the language is made of (Varela 1991).

Contrarily, embodied theories (Barsalou 2008, 2016; Johnson 1987; Lakoff 1987; Lakoff and Johnson 1980, 1999; Wilson 2002) argue that reasoning, concepts, and language are grounded in experience and tightly bonded to the body and its specific features. In this framework, it is claimed that the body and its inherent way of functioning and interacting in the physical space, directly impinge on our cognitive structures, and it thus represents one of the primary sources in the operation of meaning construction (Lakoff and Johnson 1999). A consistent number of empirical studies indeed showed that conceptual knowledge is deeply rooted in perceptual and motor systems (Gallese and Lakoff 2005; Martin and Chao 2001; Pulvermüller 2005). Additionally, it was shown that sensory-motor simulations directly impinge on the processing and understanding of language (Glenberg and Kaschak 2002; Tettamanti 2005).

The adoption of an embodied approach to the study of lexicon relies on the idea that bodily properties have a crucial function in meaning construction processes. Embodied theories, in fact, directly look at body and language as a tight coupling, in which the comprehension of the latter cannot take place without information deriving from the former (Gibbs and Colston 1995; Gibbs 2005). As I will show, bodily features, sensory inputs, and action-oriented schemas do also play a pivotal role in the construction and extension of the action verbs' meaning, both on the concrete and the abstract representation level (Panunzi and Vernillo 2019). This is why, in this paper, not only physical but also figurative meanings of action verbs have been accounted for by working on the idea that sensory-motor processes can provide us with more data on human understanding and representation of concrete and abstract concepts. The starting point of the analysis will be that the different semantic layers (i.e., primary and marked variation) characterizing the semantic core of action verbs should not be viewed as separate dimensions of the lexical meaning but, rather, as deeply and strongly connected.

3.2 Conceptual Metaphor Theory

The Conceptual metaphor theory (henceforth CMT: Lakoff and Johnson 1980, 1999) represents one of the most powerful theories on abstract reasoning. Over the years, CMT has benefited from a consistent number of empirical researches which guaranteed, in some way, the reliability of the approach (Casasanto and Bottini 2014; Gibbs 2006; Jamrozik et al. 2016). One of the essential claims of CMT is that metaphors concern not only the way we use language but also the way we organize human thought. In this theoretical scenario, metaphors are not conceived as mere rhetoric tropes but rather as cognitive processes, by means of which aspects of human cognition, perception, and experience are transposed in language (Lakoff and Johnson 1980). CMT can be considered as the most embodied approach to the study of language. It is in fact based on the idea that the way we refer to abstract

concepts exploits the rich flow of information which we gain from our experience of the world and of the way we bodily interact with the world and the objects therein. According to Lakoff and Johnson (1980: 115), a large number of concepts that are meaningful to us are either abstract or not well delineated in our experience. They thus necessitate being conceived via concrete concepts that we can understand in clearer terms. The internal structure of many abstract concepts, such as Changes, States or Causes, appears to be cognitively grounded in the metaphorical mapping of more concrete schemas as, say, FORCE and MOTION (Gibbs 2006; Lakoff and Johnson 1980, 1999). People talk about state changes in the same way they talk about motion changes (e.g., CHANGE OF STATE IS CHANGE OF MOTION), causes in the same way as forces (e.g., CAUSES ARE FORCES), or states as physical locations in the space (e.g., STATES ARE LOCATIONS).

Metaphors are based on a conceptual mapping operation that transfers preconceptual knowledge from one concrete source domain to an abstract target domain (Lakoff and Johnson 1980). The information transfer must respect some basic rules and is supposed to be constrained by a number of different factors that can enable or stop the metaphorization process (Brygida Rudzka-Ostyn 1995). Amongst others, it is worth noticing that the mapping is not an exhaustive process, that is, not all but just some aspects of the source domain are transferred onto the target domain (see *partial metaphorical utilization* phenomenon in Kövecses 2010). The mapping is conditioned by an asymmetrical directionality, according to which the transfer may only go from the source to the target domain and not vice-versa (Lakoff and Johnson 1980). Moreover, the mapping operation must not violate the internal structure of the target domain (i.e., *target domain override*). According to the Invariance Principle Hypothesis (Lakoff and Turner 1989; Lakoff 1990, 1993; Turner 1991), the metaphorical mapping must preserve the cognitive topology (or image-schematic structure) of the source domain consistently with the inherent structure of the target domain.

As the present work is concerned with the analysis of the semantic variation of action verbs, both on the concrete and the abstract level of representation, an approach to the study of language, such as proposed by the CMT, can help: (a) To better disclose the nature of the relationship that seems to tie up together the primary and metaphorical uses of a given action verb; and (b) to investigate the specific role that bodily-action information plays within our conceptual system (Panunzi and Vernillo 2019).

3.3 Image Schema Theory

Image schemas (or schemata) are a key notion in the field of Cognitive Linguistics used to tie up together embodied experience, cognition, and language. The early notion of the concept dates to the empirical works on spatial relations terms by Talmy (1983) and Langacker (1987), but it has been fully developed only a decade later by Johnson (1987) and Lakoff (1987). Image schemas have been investigated not only in

Cognitive linguistics but in many research fields, amongst others, Psycholinguistics (Gibbs and Colston 1995), Developmental Studies (Mandler 1992; Mandler and Cánovas 2014), Poetics (Lakoff and Turner 1989), and Neurosciences (Feldman and Narayanan 2004; Gallese and Lakoff 2005).

Image schemas are deemed to be imaginative structures of understanding; by their means, we can make sense of our everyday bodily functioning and physical interaction within the surrounding space. They directly emerge from bodily experience and represent a sort of bridge between sensory-motor information and higher cognitive functions (Hampe 2005). According to Johnson's (1987: XIV) traditional definition, an image schema is a 'recurring, dynamic pattern of our perceptual interactions and motor programs that gives coherence and structure to our experience'. In the literature, the umbrella term *image schema* has been subject to different interpretations and has thus resulted in a large cross-linguistic variation in the use of the term itself (Mandler and Cánovas 2014; Talmy 1983). Although there is no general agreement upon the definition of the concept, there is broad consensus that image schemas are characterized by a stable set of recurrent properties (Cienki 1997; Gibbs 2006; Hampe 2005; Hampe and Grady 2005; Johnson 1987; Krzeszowski 1993; Lakoff 1987):

a. They recur across a large variety of distinct experiences and are not bound to a particular context of experience and knowledge;
b. They are preconceptual primary components, that is, unlike propositions, they do not state the truth or other conditions of satisfaction;
c. They are characterized by having an internal gestaltic configuration (they contain a small number of related parts and intended as coherent and meaningful wholes);
d. They tend to be co-experienced together (e.g., superimposition);
e. They show an orientation towards the positive or negative default evaluation when used in metaphorical mappings (plus-minus or axiological parameter);
f. They have both a static and dynamic nature (they can represent either a state of being or processes);
g. Their internal structure of image support inferences; and
h. They operate beneath the level of our conscious awareness.

A condensed inventory of the image-schematic structures which most frequently recur in our experience is provided in Johnson (1987). The list is not conceived as a closed set but rather as the result of an informal analysis (or reflective interrogation) of the most basic phenomenological features of our every-day experience. Different approaches to the identification and categorization of image-schematic structures have been proposed by Mandler (1992), Talmy (2000), Mandler and Cánovas (2014). Beyond the differences between the various taxonomic proposals, some of the most frequently cited examples of image schemas are CONTAINMENT, SOURCE PATH GOAL, VERTICAL AXIS, FORCE, SUPPORT.

Image schemas are operative in our perceptual interactions, bodily movements, and physical manipulation of objects since early infancy (Mandler 1992; Mandler and Cánovas 2014). They are recognized as primitive cognitive components in the

development of human thought. These conceptual building blocks encode not only spatial[3] and bodily related information but also play an essential role in the modeling of highly abstract concepts (e.g., *over* in Lakoff 1987 and Brugman 1988; *verticality* in Ekberg 1995; *straight* in Cienki 1998; *smooth-rough* in Rohrer 2006). Skeletal projections of image schemas are transferred from domain to domain through analogical reasoning and metaphorical mapping (Kövecses 2010). In the operation of metaphorical mapping, image schemas constrain the information transfer in such a way to prevent the source domain topology from incoherence or inconsistency with the internal structure of the target domain (Invariance principle; Lakoff 1990, 1993; Turner 1991).

Since this linguistic investigation rests on the basic idea that physical experiences can be thought of as one of the most important sources that give meaning to conceptual structures, my analysis strongly benefited from the adoption of an image-schematic approach to the study of language. As I will show in the next paragraphs, the differential image-schematic structures characterizing the semantic core of action verbs strictly impinge on their extension and, consequently, their metaphorical potential. They determine the type of abstract concepts (and figurative meanings) that may or may not be conveyed by the action predicates. Against this background, the detection of the image schemas operating within the primary variation of action verbs helped on two levels of the analysis: first, image schemas may be used to motivate the synonymousness relations between two action verbs (e.g., both *spingere* and *premere* may be used to express the same action concept); and second, to understand the divergent or convergent behaviors that two action verbs have when used to encode abstract concepts and figurative meanings (e.g., the verbs *spingere* and *premere* are not always used to convey the same kind of metaphorical concepts).

4 Data and Methods

This study aims at investigating the semantic variation of a cohesive group of four action verbs that, in their basic meaning, codify the exertion of physical force: *premere* (Eng., *to press*), *spingere* (Eng., *to push*), *tirare* (Eng., *to pull*) and *trascinare* (Eng., *to drag*). The data the analysis was built upon have been primarily extracted from the corpus IMAGACT (Moneglia 2014a, b) and later integrated with a larger number of occurrences taken from the Opus corpus (Italian subtitles). The annotation process started with the scrutiny of more than 5000 occurrences, of which about 1000 were derived from IMAGACT and around 4000 from the Opus corpus. Interestingly, just a small part of the whole collection became the classification core. This means that the analysis of the metaphorical production (i.e., marked variation) of the four action verbs was only based on 300 metaphorical occurrences.

---

[3]According to Mandler (1992), Mandler and Cánovas (2014) spatial inputs are recoded in the form of image schemas during processes of perceptual meaning analysis and used as primitive conceptual components in the development of human thought.

With regard to the deep annotation process, it can be spelled out in the following three crucial steps:

1. Overall evaluation of the primary variation of each action verb with the extraction of the significant semantic properties with a strong distinctive value (e.g., differences in motor schemas, spatial relations, type of object involved, action participants);
2. Examination of the metaphorical uses of each action verb and isolation of the metaphorical conceptual structures found within their marked variation (Lakoff et al. 1991; MetaNet: Dodge et al. 2013);
3. Identification of the most salient image-schematic components (Johnson 1987; Lakoff 1987; Mandler and Cánovas 2014) for each verb, with respect to its primary and marked variation.

5 Description of the Primary Variation of the Four Action Verbs

The four general action verbs *premere* (Eng., *to press*), *spingere* (Eng., *to push*), *tirare* (Eng., *to pull*), and *trascinare* (Eng., *to drag*) can be looked at as a cohesive semantic class, in which the category of force-dynamics represent the main actor. They are, in fact, all used to express the exertion of some kind of physical force on the agent's body, animate theme, or tangible object. To simplify the representation of their semantic variations and the isolation of the common and differential traits, the presentation has been organized by coupling these verbs in 2 sub-groups: (1) one group represented by *premere* and *spingere*; (2) the other group represented by *tirare* and *trascinare*.

In Sect. 5.1, I will describe the primary variation of *premere* and *spingere*, highlighting convergent and divergent points along their axis of variation. In Sect. 5.2, I will focus on the description of *tirare* and *trascinare*, and I will try to illustrate their semantic similarities and differences, when their physical (and concrete) uses are considered.

5.1 The Primary Variation of the Verbs Premere and Spingere

As locally equivalent verbs, *premere* (Eng., *to press*) and *spingere* (Eng., *to push*) share a common sub-set of action concepts. They are applied in a small range of linguistic contexts to encode action events in which an agent interacts with an entity by exerting force on it. Interestingly, the entity is not deeply or permanently physically affected by the force and, overall, is not moved from one place to another. Both the verbs, for instance, are employed in the depiction of action events in which the force can result in: (a) An activation of the device connected to the affected entity ("Spingere/premere il pulsante"; Eng., "To push/To press the button"); and (b) the

establishment of new relations between two or more entities ("Spingere/premere il coperchio sulla scatola"; Eng., "To push/To press the lid on the box").

5.1.1 The Primary Variation of the Verb *Premere*

The equivalence of the verbs *premere* (Eng., *to press*) and *spingere* (Eng., *to push*) is not absolute and their variations do not tend to systematically converge. Besides the uses presented above, the verb *premere* also appears to codify action concepts in which the application of force on a specific entity (in the form of physical pressure) results in a mere physical manipulation (e.g., "Il fisioterapista preme sulla schiena di Maria"; Eng., "The physical therapist presses on Mary's back"). With regard to its inherent image-schematic structure, the verb *premere* bases on the FORCE schema and, unlike *spingere*, never entails the MOTION schema. The verb *premere* is mainly used to profile static scenarios, that is, to highlight the mere interaction between a force and the entity affected by the exertion of the force. Given the prototypical action imagery associated with *premere*, the image-schematic components which appear to play a relevant role in its primary variation are: COMPULSION FORCE, CONTACT, OBJECT, and BLOCKAGE.

5.1.2 The Primary Variation of the Verb *Spingere*

The verb *spingere* (Eng., *to push*) commonly expresses action events in which the exertion of force on a concrete entity has the motion as direct entailment. The motion can either be instantiated by an external force (e.g., CAUSED MOTION: "Spingere il carrello"; Eng., "To push the cart down the hall") or be spontaneous and not brought about by another force (e.g., SELF- PROPELLED MOTION: "Il nuotatore si spinge con le gambe"; Eng., "The swimmer pushes himself off of the wall"). Moreover, motion can be continuous and controlled by the agent along the overall path (e.g., CAUSED JOINT MOTION schema); or it can be discrete and controlled by the agent only in the initial phase of the event (e.g., CAUSED MOTION schema). The former MOTION schema plays a central role in the construal of those action events in which the agent has control of the theme throughout the motion (e.g., "Spingere il carrello"; Eng., "To push the chart down the hall"). The latter schema is determinant in those action events in which the agent does not experience the overall motion of the theme, and in which the motion results in a different spatial agent-theme configuration, such as in an increase of the physical distance between the agent and the entity affected by the force (e.g., "Spingere la scatola"; Eng., "To push the box away"). As the verb structure suggests, the tight association between the FORCE and the MOTION schemas is a distinctive feature of the semantic core of *spingere*. Rather than being used to encode events of mere force exertion, the verb *spingere* is mainly used in the encoding of kinetic events, that is, in events involving the shift of the location of the affected entity (animate or inanimate). As the prototypical action imagery associated with *spingere* suggests, the image-schematic components which do play a relevant role

Table 2 Differential image schemas in the variation of premere and spingere

Differential image schemas					
	Contact	Compulsion force	Motion	Path	Object
Premere	+	+	−	−	+
Spingere	+	+	+	+	+/−

in the verb primary variation are: COMPULSION FORCE, CONTACT, OBJECT, PATH, and SELF/CAUSED MOTION.

To give a general overview on the image schemas that I discussed so far, in the table below, I present a brief resume of the different components involved in the semantic core of the action verbs *premere* (Eng., *to press*) and *spingere* (Eng., *to push*), and I distinguish between salient (+), absent (−), and optional schemas (+/) (Table 2).

5.2 The Primary Variation of Tirare and Trascinare

When we use *tirare* (Eng., *to pull*) and *trascinare* (Eng., *to drag*) as locally equivalent verbs, we probably want to refer to action events in which an agent exerts a physical force (COMPULSION FORCE schema) on a theme (either animate or inanimate), such as to forcefully and roughly move it along a surface (CAUSED JOINT MOTION schema).[4] The force can be either directly applied on the affected entity (e.g., "Fabio tira/trascina il sacco della spazzatura"; Eng., "Fabio pulls/drags the garbage") or be indirectly applied using an intermediary instrument ("Giovanni tira/trascina la barca con l'argano"; Eng., "John pulls/drags the boat onto the beach with the winch"). The transfer of the object (e.g., theme) on the terrain does not happen smoothly, but it encounters some difficulties which slow down the motion of both entities which are involved (e.g., the agent and the theme). The slowing down may be caused by either the fact that the theme has a weight that impedes its motion or by the theme's reluctance to move along the path (BLOCKAGE schema). Either way, the verbs *tirare* and *trascinare* profile an action scene in which, at each step of the motion, the agent tries to forcefully overcome the resistance produced by the friction between the theme and its path along which the theme moves (RESTRAINT REMOVAL schema).

5.2.1 The Primary Variation of the Verb *Tirare*

The verbs *tirare* (Eng., *to pull*) and *trascinare* (Eng., *to drag*) are tied up in a relationship of partial synonymy, that is, they are not always applicable in the same linguistic contexts. The semantics of the verb *tirare* is based on a larger array of action events

---

[4]The agent has control of the theme throughout the motion and not only in the beginning phase of the force-action event.

Table 3 Differential image schemas in the variation of tirare and trascinare

Differential image schemas

	Contact	Compulsion force	Restraint removal	Blockage	Motion	Path	Object	Surface
Tirare	+	+	+/−	+/−	+/−	+/−	+	+/−
Trascinare	+	+	+	+	+	+	+/−	+

and schemas. In general, the predicate describes action scenes in which the force applied may or may not result in events of proper motion. In cases where it does, the predicate describes events in which an agent causes an object to move along a path (e.g., CAUSED MOTION[5]). The motion can be performed either along the vertical or the horizontal axis, and it is normally supposed to be directed towards the agent or towards the effector who applied the force. In cases in which the exertion of force does not result in a schema of motion, the predicate is used to profile action events involving the mere manipulation or modification of the shape of an object (e.g., "Mario tira la corda"; Eng., "Mario pulls the rope"). Given the prototypical action imagery associated with the verb *tirare*, the following image-schematic components were isolated: COMPULSION FORCE, OBJECT, CONTACT, PATH, and CAUSED/CAUSED JOINT MOTION.

5.2.2 The Primary Variation of the Verb *Trascinare*

The verb *trascinare* (Eng., *to drag*) has a primary variation narrower than that of the verb *tirare*, as it is only used to encode action events in which the motion is performed in the same agent or effector's direction (CAUSED JOINT MOTION schema). The verb *trascinare* can also be used to name physical events of SELF- PROPELLED MOTION, that is, events in which an animate entity moves along a path spontaneously, without the intervention of an external force (e.g., "Fabio si trascina lungo il corridoio"; Eng., "Fabio drags himself along the ground"). In both the cases (CAUSED and SELF- PROPELLED MOTION schemas), the predicate encodes action events in which the existence of a frictional force influences the specific manner of motion along the path (the motion is performed forcefully and roughly). As the analysis of the action imagery associated with the verb *trascinare* suggests, the following image schemas are relevant within its semantic core: COMPULSION FORCE, OBJECT, CONTACT, PATH, SELF/CAUSED JOINT MOTION, SURFACE, BLOCKAGE, RESTRAINT REMOVAL.

The following table proposes a set of differential image-schematic components that allow the better understanding of the application conditions of *tirare* (Eng., *to pull*) and *trascinare* (Eng., *to drag*) and I distinguish between salient (+), absent (−), and optional schemas (+/−) (Table 3).

---

[5]Unlike *trascinare,* the verb *tirare* does not encode the image schema SELF- PROPELLED MOTION.

6 Description of the Marked Variation of the Four Action Verbs

In the previous Sections, it has been claimed that the semantics of general action verbs is strongly tied to specific perceptual, spatial, and motor schemas. It has been shown that the semantic variation of two similar action verbs (e.g., *premere* and *spingere*; *tirare* and *trascinare*) can partially converge and be responsible for their mutual use in the operation of action reference and labeling. However, it has also been pointed out that these same verbs can also be applied in diverse pragmatic contexts to express diverse types of action events. The question I want to investigate is whether these couplings extend the same kind of interwoven semantic relations to their marked variations. Their pervasiveness, though, not manifests itself only on the level of the reference to concrete actions, but also on a more abstract one, where the semantic core is exploited to encode figurative meanings (i.e., marked variation), springing from largely metaphorical processes.

In the following Sections, it will be shown how different semantic properties of the predicates connect to a different type of metaphors and metaphorical meanings. In particular, in Sect. 6.1, the most significant types of metaphors detected within the marked variation of *premere* (Eng., *to press*) and *spingere* (Eng., *to push*) will be analyzed and compared. Finally, in Sect. 6.2, the metaphorical uses of *tirare* (Eng., *to pull*) and *trascinare* (Eng., *to drag*) will be spelled out. The analysis will not only consider the conceptual metaphorical structures needed to explain the array of abstract uses identified in the verb's semantics, but it will also identify the image schemas that are salient in the operation of metaphorical meaning construction.

6.1 The Marked Variation of the Verbs Premere and Spingere

It often happens that the verbs *premere* (Eng., *to press*) and *spingere* (Eng., *to push*) are co-extensively used to linguistically express the same kind of metaphorical concepts. Both the verbs are involved in the encoding of the general conceptual metaphor PSYCHOLOGICAL FORCES ARE PHYSICAL FORCES, via which psychological manipulation (e.g., impact or influence) is understood in terms of physical manipulation (e.g., contact or pressure):

(9)   "L'oratore *preme* sui temi sociali"
              "The speaker is pressing on social agenda"
(10)   "Occorre *premere* sulle due parti perché il negoziato sia vero"
              "We need to put pressure on the parties to make the agreement true"
(11)   "Bisogna *spingere* sui processi di liberalizzazione"
              "We need to put pressure on the deregulation processes"
(12)   "Abbiamo *spinto* affinché tale diritto sia reso più accessibile"
              "We pushed to make this right more accessible"

The verbal items in (9–12) exploit our knowledge of the category of force dynamics in the representation of the psychological interaction between two entities: the source of the force (ANIMATE ENTITY schema) and the party affected by the force (OBJECT schema). The sentences (9–12) represent, on the level of language, the projection of the abstract domain of psychological forces (e.g., influence) into the concrete domain of physical forces (e.g., pressure).

6.1.1 The Marked Variation of the Verb *Premere*

Unlike *spingere* (Eng., *to push)*, the verb *premere* (Eng., *to press*) is often used to describe a situation in which the entity exerting the force is perceived as a burdensome object (OBJECT schema), weighing on another entity or theme (OBJECT and SUPPORT schema) through a sort of imagery contact:

(13)   "La disoccupazione *preme* sulla spesa sociale"
         "Unemployment weighs on public expenditure"

Example (13) is a linguistic variation of the metaphor IMPEDIMENTS TO IMPROVING ECONOMIC STATUS IS PHYSICAL BURDEN which represents a complex case of the primary metaphorical structure DIFFICULTIES ARE IMPEDIMENTS TO MOVEMENT. The sentence frames a very specific scene in which unemployment (OBJECT schema) is conceived as a social burden or as an obstacle (BLOCKAGE schema) that weighs on (COMPULSION FORCE schema) the public spending. More in general, the verb *premere* appears to be pervasively used in the picturing of metaphors that exploit our experience of and response to burdens and loads to structure more highly abstract domains. In the same way that when I say that "Il tempo preme" (Eng., "Time is pressuring me"), I am not referring to the fact that I may eventually change the situation in which I am because of the time pressure. I am focusing on the fact that another entity (e.g., time) is exerting a psychological force (conceived in terms of pressure), that the same entity is affecting my state of mind, and that I may be weighed down by the force itself. In similar cases, the direct contact between the source and the target entity does result in a sort of burdensome stasis or mere physical pressure, without implying a change of state or action of the target entity. This fact can be connected to the fact that, as I said in (4.1.1), the action imagery associated with the verb *premere* does not entail the image schema of MOTION. As a consequence, this action verb is mainly used to represent static scenarios, that is, to express the mere interaction between a force and the entity affected by the force.

6.1.2 The Marked Variation of the Verb *Spingere*

The verb *spingere* (Eng., *to push*) rather appears in contexts where the encoding of more dynamic metaphorical concepts is based on the source domain of MOTION:

(14)   "Le circostanze *spingono* Fabio ad agire"
         "Circumstances are pushing Fabio to act"
(15)   "La situazione *si spinge* verso l'anno successivo"
         "The situation will press on into the next year"
(16)   "L'amministratore *spinge* avanti l'azienda"
         "The manager pushes the company forward"

The metaphorical extensions presented above (14–16) conceptualize causation in terms of motion (either caused or self-initiated). In example (14), external forces (e.g., circumstances) are intended in terms of animacy (ANIMATE ENTITY schema) and cause (COMPULSION FORCE schema) that a second target entity (e.g., Fabio: ANIMATE ENTITY schema) performs an action or adopt a set of actions and, eventually, behaviors (e.g., CAUSED MOTION schema). Importantly, this example bases on the generalization that caused change of action is conceived as forced motion relative to a location. The expression in (14) can be seen as the linguistic reflection of the complex conceptual metaphor CAUSED CHANGE OF ACTION IS CONTROL OVER AN ENTITY RELATIVE TO A LOCATION, which is an entailment of the metaphor CHANGE OF ACTION IS CHANGE OF MOTION. This conceptual structure also makes use of the metaphors CAUSES ARE FORCES and CAUSATION IS OBJECT TRANSFER. In example (15), an event is seen as a moving entity (ANIMATE ENTITY schema) directed from one location in space (SOURCE POINT FOCUS schema) to another (END POINT FOCUS schema). The changing that the event undergoes is understood as self-initiated motion (SELF-PROPELLED MOTION schema). The example (15) is a linguistic variant of the metaphor THE PROGRESS OF EXTERNAL EVENT IS A FORWARD MOTION,[6] but may also be understood in a more general metaphorical scenario in which TIME is conceptualized as a LANDSCAPE WE MOVE THROUGH and ACTION is conceived as SELF-PROPELLED MOTION.[7] Finally, example (16) can be connected to the conceptual metaphor CONTROL OVER ACTION IS CONTROL OVER MOTION, which is a special subcase of the conceptual metaphor PURPOSEFUL ACTION IS DIRECTED MOTION TO A DESTINATION (CAUSED JOINT MOTION schema). This metaphor also entails the metaphorical structure PROGRESS IS FORWARD MOTION ALONG THE PATH. In this and in example (14), causation is intended in terms of forced motion relative to a region or a path. The main difference is the fact that the metaphorical extension in (16) bases on action imagery slightly different from the one found in (14). In this last example, the verb *spingere* does not only encode forced motion (CAUSED MOTION schema) but also the idea that forced motion is controlled along the overall path (CAUSED JOINT MOTION schema). An animate and forceful entity (e.g., the manager) may have a specific purpose (e.g., the development of the company) and may want to guide the target entity (OBJECT schema) that she controls (e.g., company) toward the final goal of the long-term, purposeful action she is bringing about (END POINT FOCUS schema).

---

[6]This metaphor is an entailment of PROGRESS IS FORWARD MOTION ALONG THE PATH.

[7]It is a subcase of the metaphor ACTION IS MOTION ALONG THE PATH.

The combination of the FORCE and MOTION schema is also salient in the encoding of the orientation metaphorical extensions by the verb *spingere*. This is in those uses in which the predicate expresses the change of a certain value along a measurable scale:

(17)   "Tali fattori hanno spinto verso l'alto i prezzi"
            "These factors pushed up the prices"
(18)   "I rincari hanno spinto l'inflazione verso valori superiori al 2%"
            "Price increase pushed the inflation over 2%"

Both cases (17–18) can be linked to the metaphor CAUSE INCREASE IN QUANTITY IS CAUSE UPWARD MOTION, entailment of the more general primary metaphor MORE IS UP, and of the metaphor CAUSED CHANGE OF STATE IS CAUSED CHANGE OF LOCATION. The metaphorical mapping is built upon image-schematic knowledge: while the target domain (e.g., QUANTITY) makes use of the SCALE schema, the source domain (e.g., CAUSED UPWARD MOTION) makes use of the combination of the image schemas of COMPULSION FORCE, CAUSED MOTION and VERTICAL ORIENTATION.

Taken together, in all explained examples (14–18), there are two points especially interesting for my analysis: first, the category of force systematically intersects with that of motion; and second, unlike *premere*, the verb *spingere* encodes this constant semantic combination in the unravelment of both its primary and metaphorical variation.

6.2   The Marked Variation of the Verbs Tirare and Trascinare

The metaphorical variation of the verbs *tirare* (Eng., *to pull*) and *trascinare* (Eng., *to drag*) usually converge to the encoding of those conceptual metaphors that construe the domain of CAUSATION on the basis of the domains of FORCE and MOTION. The two predicates are involved in the linguistic representations of a large system of metaphors in which causation is connected to animacy (e.g., CAUSATION IS AGENTIVE CAUSATION), causes are intended in terms of force (e.g., CAUSES ARE FORCES), changes of state (or of action) are conceptualized as changes of motion (e.g., CAUSATION IS CONTROL OVER AN ENTITY RELATIVE TO A LOCATION):

(19)   "Marco tira Luca nella conversazione"
            "Marco involves Luca in the conversation"
(20)   "Il governo non tirerà l'Algeria fuori dal solco in cui si trova"
            "The government will not get Algeria out of its current situation"
(21)   "Ci hai trascinato in mezzo ai guai"
            "You dragged us in a lot of problems"
(22)   "Il presidente ha trascinato il paese sul fondo"
            "The president dragged the country down"

In examples (19–22), the verbs *tirare* and *trascinare* are used to depict metaphorical scenes in which the change of state of the affected entity is caused by an external entity (ANIMATE ENTITY schema). The agent has control over the whole process of transition from a state to another (PATH schema), and causes (COMPULSION FORCE schema) that the final state or goal achieved by the affected entity is intended in terms of motion from one location to another (CAUSED JOINT MOTION schema).[8] As the analysis of the examples shows, there exists an evident correspondence between the metaphorical extensions of the verbs *tirare* and *trascinare*, and the specific sensory-motor imagery associated with these same predicates. All the metaphorical items discussed above (19–22) are built upon an operation of conceptual mapping in which:

a. The agent corresponds to the agent that leads the motion;
b. The party affected by the new situation or process corresponds to the entity (animate or inanimate) moved by the agent along the path;
c. The caused change of state or situation corresponds to the motion caused by the agent;
d. The achievement of the final goal corresponds to the reach of the final location along the path.

6.2.1   The Marked Variation of the Verb *Trascinare*

The metaphorical variation of the verb *trascinare* (Eng., *to drag*) diverges from that of *tirare* in many points. The systematic combination of the FORCE and MOTION schemas stands as the thread that deeply connects the sets of different metaphorical uses produced by the verb. Nevertheless, either the MOTION and the FORCE schemas (and imageries) associated with the predicate are richer and more complex than those involved in the variation of *tirare,* as they seem to be more semantically constrained. Unlike *tirare*, the verb *trascinare* does not simply encode the schema of CAUSED MOTION but also that of SELF- PROPELLED MOTION. The verb does also require a specific manner of motion (frictional,[9] forceful, and difficult). With regard to the FORCE schema, the verb *trascinare* requires that the target entity is reluctant or difficult to move (BLOCKAGE schema) and that the force moving the target entity (COMPULSION FORCE schema) tries to continuously overcome that physical restraint (RESTRAINT REMOVAL schema). The metaphorical items identified within the variation of *trascinare* confirm the saliency of all the semantic aspects discussed above (see also Sect. 5.2). The CAUSED MOTION image schema seems to play a structural role within the modeling of many metaphorical uses:

---

[8]The change of state can be enriched with additional space information and represented as a motion performed along a bounded path (CONTAINER schema) or along the vertical axis (VERTICAL AXIS schema).

[9]The verb *trascinare* (Eng., *to drag*) always implies a sort of friction between the moving entity and the ground along which the entity moves.

(23)   "Gli eventi trascinano la massa"
            "The events are dragging people [along]"
(24)   "L'odio ti trascina"
            "Hatred tugs on you"
(25)   "L'attore trascina il pubblico"
            "The actor drags the audience along"
(26)   "Il tifo non trascina nessuno"
            "The cheer does not grab [lit., drag] anyone"

Sentences (23–26) profile an extremely unbalanced system of forces, in which one entity (an agent, external event, process, or emotion) is conceptualized in terms of volition and animacy, and impinges on a second entity's behavior, state, or action. The general conceptual metaphors to which we can relate these examples are the same as cited in the previous Section (e.g., CAUSATION IS AGENTIVE CAUSATION, CAUSES ARE FORCES, and CAUSATION IS CONTROL OVER AN ENTITY RELATIVE TO A LOCATION). What happens to be very interesting here (23–26) is that the verb *tirare* cannot be applied in these same metaphorical contexts to express the same kind of metaphorical meaning. The kind of force encoded by *tirare* does not happen to entail the same state of unbalance (and of the unbalanced ratio between the entities and the forces involved) that seems to be a salient feature at the base of all the metaphorical extensions expressed by *trascinare*. Unlike *tirare*, the verb *trascinare* always entails the existence of a sort of impediment to motion and, hence, the presence of a specific bodily response to that same impediment: the verb *trascinare* entails that the motion (and the action) is performed with difficulty and that difficulty increases the effort needed to accomplish an objective or to reach a goal (e.g., conceptual metaphor DIFFICULTIES ARE IMPEDIMENTS TO MOVEMENT).[10] For the same reason, the verb *trascinare* is mainly used to encode metaphors that imply a slightly negative meaning. The same characteristics discussed so far seem to be relevant to the metaphorical extensions of the verb *trascinare* that rely upon a different type of motion schema, that is, the self-propelled motion schema. In the case of self-propelled motion, instead of being affected by an external force, one entity moves spontaneously with its direction:

(27)   "Il conflitto si trascina da anni"
            "The war drags on for years"
(28)   "Gianni si trascina in un'esistenza spaventosa"
            "Gianni is dragging himself into an awful existence"

Examples in (27–28) have different meanings and refer to different abstract concepts, but both can be linked to the primary conceptual metaphor SELF-PROPELLED ACTION IS SELF-PROPELLED MOTION. While in the first sentence (27) the moving entity is represented by a long-lasting event (e.g., TIME IS A LANDSCAPE IN WHICH EVENTS

---

[10]For the same reason, the verb *trascinare* (Eng., *to drag*) is mainly used to encode metaphors that imply a slightly negative meaning (see plus-minus parameter in Krzeszowski 1993).

MOVE THROUGH),[11] in the second example (28) the moving entity is represented by a person, a volitional, and animate entity, who laboriously drags herself in a painful and difficult situation. Interestingly, in (27–28), the verb *tirare* cannot be applied since it does not happen to encode, with its semantic core, the schema of SELF- PROPELLED MOTION. On the contrary, in these sentences, the verb *trascinare* is perfectly usable since it also codifies the SELF- MOTION schema in its primary variation.

6.2.2 The Marked Variation of the Verb *Tirare*

As we saw in examples (19–20), the verb *tirare* (Eng., *to pull*) is mostly used to encode causation events, that is, to profile metaphorical scenarios in which one entity causes another entity to be affected by the occurrence of a new event or state (e.g., CONTROL OVER ACTION IS CONTROL OVER MOTION, CAUSED CHANGE OF STATE (or ACTION) IS CAUSED CHANGE OF MOTION, etc.). Interestingly, this verb often encodes causation events which entail a specific spatial relationship between the agentive force and the entity affected by the force:

(29)   "Non hai speranze di tirarmi dalla tua parte"
             "You cannot get me on your side"
(30)   "Sandra tira sempre"
             "Sandra is attractive"

Metaphors in (29–30) show that the PATH schema involved in the semantic core of *tirare* entails that the shift from point A (start point focus schema) to point B (END POINT FOCUS schema) which is performed by the entity affected by the force corresponds to the spatial location of the source of the force. The verb implies that the motion is directed towards the actor, that is, towards the source of the force (TOWARDS TO schema; NEAR FAR schema). More in particular, the example (29) is a subcase of the metaphorical structure AGREEMENT IS BEING ON THE SAME SIDE (OR AGREEMENT IS PROXIMITY), in which physical closeness is the source domain for metaphors of similarity, solidarity, and support. The example (30) may be seen as a linguistical extension of the conceptual metaphors DESIRES THAT CONTROL ACTION ARE EXTERNAL FORCES THAT CONTROL MOTION[12] and DESIRES ARE FORCES BETWEEN THE DESIRED AND THE DESIRER. Thereby sexual attraction is interpreted as a force toward physical proximity or closeness (e.g., ATTRACTION FORCE schema), and the desired object is interpreted as a desired state or location. The verb *trascinare* cannot be applied in similar metaphorical contexts, for two main reasons: first, its action-motion schema presupposes that both the agent and

---

[11]Interestingly, when the moving entity is represented by an inanimate entity, the verb *trascinare* always encodes figurative meanings in which the duration of a process (event or situation) is measured in terms of motion along a path.

[12]This metaphor also could be associated with example (25). Nevertheless, the verb *trascinare* does not bring along the same kind of inferential structure as *tirare* and does not entail that the attraction force between the agent and the target entity results in a different spatial configuration between the two.

the affected entity are in motion (CAUSED JOINT MOTION schema); and second, even though they move in the same direction (the agent's direction), the final point reached by the affected entity does not correspond to the agent's location and does not result in a sort of shortening of the distances between the entities (TOWARDS TO schema; NEAR FAR schema). The verb *tirare* seems to be also pervasively used in the encoding of orientational metaphors, that is, metaphors whose mapping organizes target concepts by means of very basic spatial vectors, such as up-down, near-far, in–out, center-periphery, and so on:

(31)   "L'insegnante tira su il voto di Luca"
            "The teacher raises Luca's grade"[13]

(32)   "Ho provato a tirarlo su"
            "I tried to cheer him up"

In the example (31), the PATH schema is conceived as a SCALE, i.e., as a vertical path, whose points are not intended as neutral points but as values. It profiles a scenario in which an actor (ANIMATE ENTITY) causes an entity (OBJECT schema) to change position on a scale. The change of position from a point (START POINT FOCUS schema) to another (END POINT FOCUS schema) results in a change of state of the object (here conceived as a value). The metaphorical extension in (31) can be interpreted as a lexical representation of the metaphor CAUSE INCREASE IN QUANTITY IS CAUSE UPWARD MOTION, which is a special case of the more general and primary conceptual metaphor MORE IS UP. Finally, example (32) represents a scenario in which the passage from a negative to a positive emotional state is conceptualized in terms of upward motion, this is caused by an external force or entity. The expression is a case of the conceptual metaphor CAUSE CHANGE IN MOOD IS VERTICAL MOTION, which is a subcase of the primary metaphor HAPPY IS UP (or IMPROVEMENT IN MOOD IS UPWARD MOTION).

7   Discussion of the Results

This work focused on the semantic description of four action verbs encoding force, i.e., *premere* (Eng., *to press*), *spingere* (Eng., *to push*), *tirare* (Eng., *to pull*), and *trascinare* (Eng., *to drag*). The analysis was organized in a way to simultaneously compare two pairings of verbs: on the one hand, similarities and differences between the verbs *premere* and *spingere* were presented; on the other hand, convergences and divergences between *tirare* and *trascinare* were explained.

---

[13]The action verb *trascinare* (Eng., *to drag*) cannot be applied to encode the metaphorical increase (or decrease) of a value along an imagery VERTICAL AXIS (SCALE schema). This predicate can only be used to encode force-motion events along the horizontal axis. The kind of force encoded by *trascinare*, in fact, presupposes that the gravitational steady state of the entities involved in the event does not change. The entities must move along the ground (or horizontal path), producing a continuous frictional force.

With regard to the primary variation, it has been shown that the action verb *premere* (Eng., *to press*) only applies to contexts in which the state of the theme affected by the force does not result in any form of motion (BLOCKAGE schema). Additionally, it has been stressed that *premere* focuses on the pure exertion of force (in the form of physical pressure), i.e., on the interaction between the entity that applies the force and the object towards which the same force is directed. Unlike *premere*, the other three verbs encode the MOTION schema within their inner semantic skeleton, thus being used to profile more kinetic action scenes. Both the verbs *spingere* (Eng., *to push*) and *tirare* (Eng., *to pull*) have very flexible semantics, being able to encode different types of action events (with or without the association of force and motion). Nevertheless, they mainly focus on the result of the forceful interaction between the entities involved in the action, that is, on the directed caused motion to which the object is subject to. In *spingere*, the motion is normally thought to be directed from the point of contact between the effector and the object and away from the agent; in *tirare*, the motion is normally thought to be directed from the point of contact between the effector and object, and towards the agent. Finally, the verb *trascinare* (Eng., *to drag*) represents a very specific case, as it requires a greater number of necessary components for its application, and always needs the CAUSED JOINT MOTION and the RESTRAINT REMOVAL schemas to be activated. As a matter of fact, in *trascinare* the application of force happens to be always associated with the MOTION schema., and it is, in some way, limited by the fact that the object has a weight and may be reluctant to move. These two facts currently represent a restraint that is going to be constantly removed to move the object along the surface it lies upon.

This study not only aimed at showing how the semantics of action words mirrors the way in which we internally structure the logic of metaphorical concepts. As it has been stressed throughout the analysis, the differential semantic traits that characterize the four predicates strictly influence their metaphorical potential. When their semantic network converges, it is easier to detect the reasons why these predicates can be equally applied to express the same figurative meanings. On the contrary, when their semantic extensions start to diverge, we may wonder how it is possible that some metaphorical concepts can be accessed by one verb and not by the other. On the basis of the data, I suggest that for a metaphor to be expressed in a specific context, the predicate must contain specific schemas pertaining to that context. With respect to the evaluation of these four action verbs semantics:

(a) Metaphors involving the target domains of PSYCHOLOGICAL MANIPULATION, INFLUENCE or IMPACT are enabled by the presence of the FORCE (exerted in the form of a pressure) image schema and have been encountered in the variation of the predicates *premere* and *spingere*;

(b) Metaphors encoding CAUSATION are enabled by the combination of the FORCE and MOTION image schemas, and have been detected only along the semantic variation of *spingere, tirare* and *trascinare*;

(c) Metaphors encoding self-propelled changes of state or action are enabled by the combination of the FORCE and SELF- PROPELLED MOTION image schemas, and have been only found in the metaphorical production of *spingere* and *trascinare*;

Table 4 Metaphorical potential of the four action verbs

Metaphor type	Image schemas	*Premere*	*Spingere*	*Tirare*	*Trascinare*
Manipulation, influence, impact	Force	+	+	-	−
Causation	Force, caused motion	−	+	+	+
Self-caused change of state or action	Force, self-propelled motion	−	+	−	+
Spatial orientation	Force, motion, vertical orientation	−	+	+	−

(d)   Orientational metaphors are enabled by the presence of the VERTICAL ORIEN-TATION image schema and have been identified only with the annotation of *spingere* and *tirare*, which happen to be less spatially constrained than, say, *trascinare.*

The following table schematizes the relationship between the verbs and their metaphorical potential (Table 4).

8   Conclusions

The data extracted from the semantic variation of the verbs *premere* (Eng., *to press*), *spingere* (Eng., *to push*), *tirare* (Eng., *to pull*) and *trascinare* (Eng., *to drag*) suggest that the metaphorical extensions of these action verbs are not randomly produced but are the result of metaphorical processes in which sensory-motor information and specific image-schematic features are transferred from one domain to another, to enable the representation of highly abstract concepts. In particular, it was shown that differential semantic properties (and image-schematic structures) characterizing the verbs strictly impinge on their metaphorical potential, determining, in some way, the type of metaphorical items that may or may not be expressed (Lakoff 1990, 1993; Turner 1991). The analysis also shows that the same differential semantic properties (and image-schematic structures) are also responsible for the type of partial equivalence that can be established between the action verbs (e.g., *premere* and *spingere*), either when their primary or marked variations are considered. In this sense, the investigation of the action verbs' semantics contributes to a better understanding of the way we use action information and very basic bodily schemas to shape not only the way we think but also the way we talk. Action verbs constitute essential linguistic anchors between sensory-motor experience and abstract knowledge, whose deeper semantic description may be used to a different number of goals and, especially, in the building up and structuring of linguistic resources and ontologies. Even in the IMAGACT ontology, a more articulated characterization of action lexicon may be used to improve the representation of verbs' senses, and to systematically define the linguistic boundaries between sense extensions of similar action verbs (e.g., locally

equivalent verbs). Finally, the image-schematic approach may be a useful tool in the representation of the metaphorical network activated by each action verb stored in the Ontology. To reframe the research on a more general level, I believe that the current results may give their main contribution to the field of Cognitive Linguistics and semantic studies.

References

Aziz-Zadeh, L., & Damasio, A. (2008). Embodied semantics for actions: Findings from functional brain imaging. *Journal of Physiology-Paris, 102*(1), 35–39.

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

Barsalou, L. W. (2016). Can cognition be reduced to action? Processes that mediate stimuli and responses make human action possible. In A. K. Engel, K. J. Friston & D. Kragic (Eds.), *The pragmatic turn: Toward action-oriented views in cognitive science*. Series: Strüngmann Forum Reports (Vol. 18). Cambridge, MA: MIT Press.

Brown, S. W. (2014). From visual prototype of action to metaphors extending the IMAGACT ontology of action to secondary meanings (ISO-10). In H. Bunt (Ed.), *Proceedings 10th Joint ISO – ACL SIGSEM, Workshop on Interoperable Semantic Annotation*. Reykjavik (pp. 53–56).

Brugman, C. (1988). *The story of over. Polysemy, semantics and the structure of the Lexicon*. New York: Garland [Originally: M.A. thesis at the University of California, Berkeley, 1981].

Casasanto, D., & Bottini, R. (2014). Mirror reading can reverse the flow of time. *Journal of Experimental Psychology, 143*(2), 473–479.

Cienki, A. (1997). Some properties and groupings of image schemas. In M. Verspoor, K. Lee & E. Sweetser (Eds.), *Lexical and Syntactical Constructions and the Construction of Meaning: Proceedings of the Bi-annual ICLA Meeting in Albuquerque July 1995* (pp. 3–15). Amsterdam/Philadelphia: John Benjamins.

Cienki, A. (1998). STRAIGHT: An image schema and its metaphorical extensions. *Cognitive Linguistics, 9*(2), 107–149.

Desai, R. H., Binder, J. R., Conant, L. L., Mano, Q. R., & Seidenberg, M. S. (2011). The neural career of sensory-motor metaphors. *Journal of Cognitive Neuroscience, 23*(9), 2376–2386.

Dodge, E., Hong, J., Stickles, E., & David, O. (2013). The MetaNet Wiki: A collaborative online resource for metaphor and image schema analysis. In *12th International Cognitive Linguistics Conference (ICLC 12)*.

Ekberg, L. (1995). The mental manipulation of the vertical axis: How to go from "up" to "out" or from "above" to "behind". In M. Verspoor, K. Dong, & E. Sweetser (eds), *Lexical and Syntactical Constructions and the Construction of Meaning: Proceedings of the Bi-annual ICLA Meeting in Albuquerque July 1995* (pp. 69–88). Amsterdam/Philadelphia: John Benjamins.

Feldman, J., & Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain and Language, 89*, 385–392.

Gagliardi, G. (2014). *Validazione dell'ontologia dell'azione IMAGACT per lo studio e la diagnosi del Mild Cognitive Impairment (MCI)*. Ph.D. Thesis. Firenze: Università degli Studi di Firenze.

Gallese, V., & Lakoff, G. (2005). The Brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology, 22*(3/4), 455–479.

Gibbs, R. W. (2005). The psychological status of image schemas. In B. Hampe (Ed.), *From perception to meaning: Image schemas in cognitive linguistics*. Berlin: Mouton de Gruyter (pp. 113–-135).

Gibbs, R. W. (2006). *Embodiment and cognitive science*. Cambridge: Cambridge University Press.

Gibbs, R. W., Jr., & Colston, H. L. (1995). The cognitive psychological reality of image schemas and their transformations. *Cognitive Linguistics, 6*, 347–378.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*, 558–565.

Hampe, B. (2005). When down is not bad, and up not good enough: A usage-based assessment of the plus–minus parameter in image-schema theory. *Cognitive Linguistics, 16*(1), 81–112.

Hampe, B., & Grady, J. (Eds.). (2005). *From perception to meaning image schemas in cognitive linguistics.* Berlin/New York: Mouton de Gruyter.

Jamrozik, A., McQuire, M., Cardillo, E. R., & Chatterjee, A. (2016). Metaphor: Bridging embodiment to abstraction. *Psychonomic Bulletin & Review, 23*, 1080–1089.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason.* Chicago: The University of Chicago Press.

Kiefer, M., & Pulvermüller, F. (2011). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex, 48*(7), 805–825.

Kövecses, Z. (2010). *Metaphor: A practical introduction.* Oxford: Oxford University Press.

Krzeszowski, T. P. (1993). The axiological parameter in preconceptual image schemata. In R. A. Geiger & B. Rudzka-Ostyn (Eds.), *Conceptualizations and mental processing in language.* (pp. 307–329). Berlin/New York: Mouton de Gruyter.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: The University of Chicago Press.

Lakoff, G. (1990). The invariance hypothesis: Is abstract reason based on image-schemas? *Cognitive Linguistics, 1*(1), 39–74.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought.* (pp. 203–249). Cambridge: Cambridge University Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Lakoff, G., & Turner, M. (1989). *More than cool reason: A field guide to poetic metaphor.* Chicago: University of Chicago Press.

Lakoff, G., Swenson, J., & Schwartz, A. (1991). *Master metaphor list* (2nd ed.). Technical report, University of California at Berkeley.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought.* New York: Basic Books.

Langacker, R. W. (1987). *Foundations of cognitive grammar, vol. 1: Theoretical prerequisites.* Stanford: Stanford University Press.

Mandler, J. M. (1992). How to build a baby: ii. Conceptual primitives. *Psychological Review, 99*, 597–604.

Mandler, J. M., & Cánovas, C. P. (2014). On defining image schemas. *Language and Cognition, 6*, 510–532.

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology, 11*, 194–201.

Moneglia, M. (2014a). Natural language ontology of action, a gap with huge consequences for natural language understanding and machine translation. In Z. Vetulani & J. Mariani (Eds.), *Human language technology challenges for computer science and linguistics* (pp. 379–395). Berlin/New York: Springer.

Moneglia, M. (2014b). The semantic variation of action verbs in multilingual spontaneous speech corpora. In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistics studies* (pp. 152–190). Amsterdam: John Benjamins.

Moneglia, M., Monachini, M., Calabrese, O., Panunzi, A., Frontini, F., Gagliardi, G., & Russo, I. (2012). The IMAGACT cross-linguistic ontology of action. A new infrastructure for natural language disambiguation. In N. Calzolari, K. Choukri, T. Declerck, M. Uğur Doğan, B. Maegaard, J. Mariani, J. Odijk & S. Piperidis (Eds.), *Proceedings of the Eight International Conference on Language Resources and Evaluation* (pp. 948–955). Paris: European Language Resources Association (ELRA).

Panunzi, A., & Moneglia, M. (2004). La Variazione Primaria del verbo nel lessico dei corpora di parlato. In F. Albano Leoni, F. Cutugno, M. Pettorino & R. Savy (Eds.), *Atti del Convegno Il Parlato Italiano, volume CD-ROM, C4* (pp. 1–24). Napoli: D'Aura Edizioni.

Panunzi, A., & Vernillo, P. (2019). Metaphor in action. Action verbs and abstract meaning. In M. Bolognesi & G. J. Steen (Eds.), *Perspectives on abstract concepts. Cognition, language and communication* (pp. 215–237). Amsterdam: John Benjamins.

Panunzi, A., De Felice, I., Gregori, L., Jacoviello, S., Monachini, M., Moneglia, M., & Quochi, V. (2014). Translating action verbs using a dictionary of images: The IMAGACT ontology. In A. Abel, C. Vettori & N. Ralli (Eds.), *Proceedings of the XVI EURALEX International Congress: The User in Focus* (pp. 1163–1170). Bolzano: EURAC Research.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience, 6*, 576–582.

Rohrer, T. (2006). Image schemata in the brain. In B. Hampe (Ed.), *From perception to meaning: Image schemas in cognitive linguistics* (pp. 165–196). Berlin: Mouton de Gruyter.

Rudzka-Ostyn, B. (1995). Metaphor, schema, invariance: The case of verbs of answering. In L. Goossens, P. Pauwels, B. Rudzka-Ostyn, J. Vanparys & A. M. Simon-Vandenbergen (Eds.), *By word of mouth: Metaphor, metonymy and linguistic action in a cognitive perspective: Studies in meaning extension in linguistic action. (Pragmatics and beyond)* (pp. 206–244). Amsterdam/Philadelphia: John Benjamins.

Talmy, L. (1983). How language structures space. In H. L. Pick & L. P. Acredolo (Eds.), *Spatial orientation: Theory, research, and application* (pp. 225–282). New York/London: Plenum Press.

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

Tettamanti, M., et al. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience, 17*(2), 273–281.

Tomasello, M. (2003). *Constructing a language.* Cambridge, MA: Harvard University Press.

Turner, M. (1991). *Reading minds: The study of English in the age of cognitive science.* Princeton, NJ: Princeton University Press.

Varela, F. J., et al. (1991). *The embodied mind: Cognitive science and human experience.* Cambridge, MA: MIT Press.

Vernillo, P. (2019). The role of the image Schemas in the analysis of the semantic variation of action verbs. Data from IMAGACT. In O. Kutz & M. Hedblom (Eds.), *Tricolore 2018.* Creativity-cognition-computation. Bozen-Bolzano, 13–15 December 2018.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*, 625–636.

Zipoli Caiani, S. (2011). The embodied theory of language: Evidence and constrains. *Logic and Philosophy of Science, 9*(1), 485–491.