

Language, Cognition, and Mind

Sebastian Löbner ·  
Thomas Gamerschlag ·  
Tobias Kalenscher · Markus Schrenk ·  
Henk Zeevat *Editors*

# Concepts, Frames and Cascades in Semantics, Cognition and Ontology

OPEN ACCESS

 Springer

# **Language, Cognition, and Mind**

## **Volume 7**

### **Editorial Board**

Tecumseh Fitch, University of Vienna, Vienna, Austria

Peter Gärdenfors, Lund University, Lund, Sweden

Bart Geurts, Radboud University, Nijmegen, The Netherlands

Noah D. Goodman, Stanford University, Stanford, USA

Robert Ladd, University of Edinburgh, Edinburgh, UK

Dan Lassiter, Stanford University, Stanford, USA

Edouard Machery, Pittsburgh University, Pittsburgh, USA

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](mailto:chungminlee55@gmail.com).

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

Sebastian Löbner · Thomas Gamerschlag ·  
Tobias Kalenscher · Markus Schrenk ·  
Henk Zeevat  
Editors

# Concepts, Frames and Cascades in Semantics, Cognition and Ontology



Springer

*Editors*

Sebastian Löbner  
Department of Linguistics  
University of Düsseldorf  
Düsseldorf, Nordrhein-Westfalen, Germany

Thomas Gamerschlag  
Department of General Linguistics  
University of Düsseldorf  
Düsseldorf, Nordrhein-Westfalen, Germany

Tobias Kalenscher  
Institute of Experimental Psychology  
Heinrich Heine Universität Düsseldorf  
Düsseldorf, Nordrhein-Westfalen, Germany

Markus Schrenk  
Heinrich Heine Universität Düsseldorf  
Düsseldorf, Nordrhein-Westfalen, Germany

Henk Zeevat  
Institute for Logic, Language  
and Computation (ILLC)  
University of Amsterdam  
Amsterdam, Noord-Holland  
The Netherlands



ISSN 2364-4109  
Language, Cognition, and Mind  
ISBN 978-3-030-50199-0  
<https://doi.org/10.1007/978-3-030-50200-3>

ISSN 2364-4117 (electronic)  
ISBN 978-3-030-50200-3 (eBook)

© The Editor(s) (if applicable) and The Author(s) 2021. This book is an open access publication.  
**Open Access** This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

This volume is a truly interdisciplinary anthology with contributions from linguistics, philosophy and psychology which cover a broad range of research on language and cognition. The articles contain theoretical, empirical and experimental work which explores the nature of mental representations that support natural language production/understanding, other manifestations of cognition as well as general reasoning about the world. Many, but not all papers in this volume, were originally presented at the conference “Cognitive Structures: Linguistic, Philosophical and Psychological Perspectives” (CoSt16) held at Heinrich Heine University Düsseldorf in September 2016. The conference, which was intended as a platform for the interchange of different perspectives on the nature of cognition, is part of a conference series. This series was realized by the Collaborative Research Centre 991 “The Structure of Representations in Language, Cognition, and Science” funded by the Deutsche Forschungsgemeinschaft DFG (German Research Foundation). Both, this book as well as the conference series, are the direct result of one of the research center’s main aims of bringing together approaches from various disciplines in order to find an adequate way for capturing aspects of concept formation in science, cognition and the description of natural language semantics.

We would like to express our gratitude to the reviewers of the single papers as well as to the two anonymous reviewers of the entire book. Without their help and insightful comments this book project would not have been possible. Furthermore, we are grateful to the DFG for the financial support of the conference series and the publication of this volume. Special thanks go to Helen van der Stelt and Anita van

der Linden-Rachmat at Springer for their experienced support and patience at any stage of the book project. Finally, we would like to thank Chungmin Lee who gave us the opportunity to publish the volume in the series “Language, Cognition and Mind”, a series we consider an ideal place for this book.

Düsseldorf, Germany  
Düsseldorf, Germany  
Düsseldorf, Germany  
Düsseldorf, Germany  
Amsterdam, The Netherlands

Sebastian Löbner  
Thomas Gamerschlag  
Tobias Kalenscher  
Markus Schrenk  
Henk Zeevat

# Contents

<b>Introduction . . . . .</b>	<b>1</b>
Sebastian Löbner, Thomas Gamerschlag, Tobias Kalenscher, Markus Schrenk, and Henk Zeevat	
<b>Pushing the Boundaries of Formal Semantics</b>	
<b>A Compositional Pluralist Semantics for Extensional and Attitude Verbs . . . . .</b>	<b>25</b>
Kristina Liefke	
<b>Counting Possible Configurations . . . . .</b>	<b>43</b>
Manfred Krifka	
<b>Structure and Ontology in Nonlocal Readings of Adjectives . . . . .</b>	<b>65</b>
Marcin Morzycki	
<b>Concept Theory</b>	
<b>How Can Semantics Avoid the Troubles with the Analytic/Synthetic Distinction? . . . . .</b>	<b>103</b>
Roberto G. de Almeida and Caitlyn Antal	
<b>Linguistic Relativity and Flexibility of Mental Representations: Color Terms in a Frame Based Analysis . . . . .</b>	<b>121</b>
Leda Berio	
<b>Implicatures and Naturalness . . . . .</b>	<b>143</b>
Igor Douven	
<b>Perception, Types and Frames . . . . .</b>	<b>165</b>
Robin Cooper	

**Conceptualizing Eventualities**

- An XMG Account of Multiplicity of Meaning in Derivation . . . . .** 181  
Marios Andreou and Simon Petitjean

- Operationalizing the Role of Context in Language Variation: The Role of *Perspective Alignment* in the Spanish Imperfective Domain . . . . .** 201  
Martín Fuchs, María Mercedes Piñango, and Ashwini Deo

- A Frame-Based Analysis of Verbal Particles in Hungarian . . . . .** 219  
Kata Balogh and Rainer Osswald

- On the Fictive Reading of German *Steigen* ‘Climb, Rise’: A Frame Account . . . . .** 239  
Thomas Gamerschlag and Wiebke Petersen

- Cascades, Goldman’s Level-Generation, Multilevel Categorization of Action, and Multilevel Verb Semantics . . . . .** 263  
Sebastian Löbner

**Prototypes and Probabilities**

- Modification and Default Inheritance . . . . .** 311  
Corina Strößner, Annika Schuster, and Gerhard Schurz

- A Frame-Theoretic Model of Bayesian Category Learning . . . . .** 329  
Samuel D. Taylor and Peter R. Sutton

- Extremes are Typical. A Game Theoretical Derivation . . . . .** 351  
Robert van Rooij and Thomas Brochhagen

- Grading Similarity . . . . .** 365  
Carla Umbach and Helmar Gust

**Cognition and Psychology**

- Escitalopram Restores Reversal Learning Impairments in Rats with Lesions of Orbital Frontal Cortex . . . . .** 389  
David S. Tait, Ellen E. Bowman, Silke Miller, Mary Dovlatyan, Connie Sanchez, and Verity J. Brown

- Rat Ultrasonic Vocalizations as Social Reinforcers—Implications for a Multilevel Model of the Cognitive Representation of Action and Rats’ Social World . . . . .** 411  
Tobias Kalenscher, Lisa-Maria Schönfeld, Sebastian Löbner, Markus Wöhr, Mireille van Berkel, Maurice-Philipp Zech, and Marijn van Wingerden

Contents	ix
<b>Influence of Manner Adverbs on Action Verb Processing . . . . .</b>	439
Jan Sieksmeyer, Anne Klepp, Valentina Niccolai, Jacqueline Metzlaff, Alfons Schnitzler, and Katja Biermann-Ruben	
<b>When Mechanical Computations Explain Better . . . . .</b>	463
Silvano Zipoli Caiani	
<b>Index . . . . .</b>	479

# Introduction



Sebastian Löbner, Thomas Gamerschlag, Tobias Kalenscher,  
Markus Schrenk, and Henk Zeevat

In order to help to explain cognition, cognitive structures are assumed to be present in the mind/brain. While the empirical investigation of such structures is the task of cognitive psychology, the other cognitive science disciplines like linguistics, philosophy and artificial intelligence have an important role in suggesting hypotheses. Researchers in these disciplines increasingly test such hypotheses by empirical means themselves. In philosophy, the traditional way of referring to such structures is via *concepts*, i.e. those mental entities by which we conceive reality and with the help of which we reason and plan. Linguists traditionally refer to the cognitive structures as *meanings*—at least those linguists with a mentalistic concept of meaning do who do not think of meaning as extra-mental entities.

The cognitive structures that are discussed in this volume are frames, conceptual spaces, prototypes, cascades, and motor representations of content. *Frames* are the attribute-value structures proposed in lexical semantics by Fillmore (1976) and in psychology by Barsalou (1992a, b). They are closely related to the attribute-value matrices in computational linguistics and knowledge representation in Artificial Intelligence. The notion of *conceptual spaces* refers to the tradition of geometrical

---

S. Löbner (✉) · T. Gamerschlag

Institute of Linguistics and Information Sciences, Heinrich-Heine-Universität, 40204 Düsseldorf, Germany

e-mail: [loebner@phil.hhu.de](mailto:loebner@phil.hhu.de)

T. Kalenscher

Comparative Psychology, Institute of Experimental Psychology, Heinrich-Heine-Universität, 40204 Düsseldorf, Germany

M. Schrenk

Department of Philosophy, Heinrich-Heine-Universität, 40204 Düsseldorf, Germany

H. Zeevat

Institute for Logic, Language and Computation, Universiteit van Amsterdam, Amsterdam, The Netherlands

approaches to meaning started with Gärdenfors (2000). *Cascades* are combinations of frames in a tree, introduced in this volume by Löbner. *Prototypes* are the idea that concepts are defined by typical cases. It is not clear that there are important divisions here. Cascades are a natural extension of frames, as they integrate several frames into a coherent more complex structure; attributes in frames often (or even always) have values in conceptual spaces and the regions defined by concepts within the spaces seem to behave much like prototypes. The *motor representations* are not accessible to introspection and can possibly be defined as frames. It seems encouraging that most of these notions can be connected to each other either by integration or by combination. There is a set of closely related hypotheses that is fine-tuned by reflection, increasing formalization, and connection to an ever widening group of phenomena.

Formal semantics does not aim directly at the cognitive level. It aims at the logical analysis of natural language, using logical relations like entailment and equivalence, and the relation between a predicate and its arguments as a probe into linguistic meaning. Meaning representations in formal semantics are essentially logical formulae for truth conditions. However, there are tendencies to take a closer look at the underlying model-theoretic ontology and provide a more differentiated landscape of things referred to. These developments provide another road of approximation to the cognitive enterprise, as the ontology relevant for natural language semantics is closely related to the way we conceive of the world. The three contributions from formal semantics by Liefke, Krifka, and Morzicky fit in here by introducing agents with a subjective epistemic perspective into the model (Liefke), and arguing for a refined ontology in the models underlying the formal interpretation of natural language (Krifka and Morzicky).

There are in principle two ways of approaching concepts: the extensional way and the intensional way. The extensional way aims at approaching concepts by getting more grip on their extensions, mostly by developing general constraints on concepts and by invoking learning. Formal semantics is the most elaborate representative of the extensional approach as it approaches conceptual meaning from outside; more on the character of formal semantics as opposed to cognitive approaches will be said in the next section. Another example is Gärdenfors' condition of convexity in a conceptual space (used in van Rooij & Brochhagen, Douven, Strößner et al.). In prototype-based accounts of concepts, one can learn a precise criterion for determining whether (or to what degree) an object falls under the concept, without thereby obtaining a conceptual decomposition that would characterize the conceptual content and thus be a properly intensional account of the content.

In a sense, all the experimental psychological contributions belong here: grounding cognitive analysis on behavioral data is an “extensional” approach, as also are approaches based on brain images: Kalenscher et al., Sieksmeyer et al. and Tait et al.

The intensional approach tries to model conceptual content. Here belong the classical approaches to concepts from Aristotle to modern cognitive theories of conceptual representation like Barsalou’s frame theory. In this volume it is all frame contributions : Berio, Andreou & Petitjean, Balogh & Osswald, Gamerschlag & Petersen, Löbner, Strößner et al., Taylor & Sutton; Cooper can also be affiliated here.

Umbach & Gust develop an original approach to similarity in which similarity ends up as a strongly context-dependent notion. This can be seen as a concern with the notion of attribute: each perspective under which a and b can be similar is in principle an attribute that applied to a and b returns identical values in some domain. It seems attributes can be made up at will—within certain limits. While this is undoubtedly an intensional approach, it also captures aspects of the geometrical way of thinking.

Several papers discuss the cognitive operations allowed by the structures. In Cooper, this is reasoning over record types with a type logic, in Löbner inferring higher levels in a cascade of frames, in Douven pragmatic reasoning in conceptual spaces. Lexical semantics has always been connected with one special cognitive operation, lexical combination to obtain the meanings of larger units than words. Learning is discussed in Sieksmeyer et al., in Tait et al. and in Taylor & Sutton.

The phenomena and approaches discussed in the papers of the volume and the fields from which they are coming span a wide area. There are philosophical discussions of enactivism (Zipoli-Caiani), the analytic-synthetic distinction (de Almeida & Antal), stereotypes (Strößner et al.), color perception (Berio), perception (Cooper), and implicature (Douven); linguistic semantic approaches to aspect (Fuchs et al.), attitude verbs (Liefke), particles (Balogh & Osswald), non-local readings of adjectives (Morzycki), derivational morphology (Andreou & Petitjean), verbs of movement (Gamerschlag & Petersen), and counting (Krifka). There are psychological studies of pragmatics and the connection between modifiers and movement (Sieksmeyer et al.), rat vocalizations (Kalenscher et al.) and rat reversal learning (Tait et al.). All approaches are relevant to the connected hypotheses mentioned above.

## 1 Cognitive Structures in Natural Language Semantics

The dominant paradigm in linguistic semantics still is the framework of formal semantics; it goes back to Richard Montague’s seminal work on the formal analysis of natural language syntax and semantics (Montague 1970, 1973). The semantic component of this framework is a model-theoretic possible-worlds semantics. Lexical and compositional meanings are essentially functions (called “intensions”) from the set of possible worlds to appropriate types of entity such as truth values (for sentences), sets of individuals in the universe (for intransitive verbs, common nouns, or one-place adjectives), or sets of sets of individuals in the universe (for quantifiers). The meaning of a sentence is given by its truth-conditions which assign, per possible world, a truth value to that sentence. The criterion of adequacy for semantic analysis is logical adequacy: do the truth-conditions account for all and only those logical entailments a sentence carries?

The approach is cast in classical Cantorian set theory. Notably, the ontology of Cantorian set theory, and consequently of mainstream mathematics, does not know things like concepts—unlike Frege’s approach to logics and mathematics. Frege distinguishes concepts and objects, (intensional) sense and (extensional) reference (Frege 1892). Montague grammar is a mathematical model of natural language grammar and meaning in this Cantorian framework. The notion of meaning is a set-theoretical, and therefore extensional *mathematical reconstruction* of Frege’s conceptual approach to linguistic meaning, notwithstanding the “conceptual” terminology introduced by Montague, who speaks, for example, of ‘intensions’, ‘properties’ and ‘individual concepts’. A central point of Montague’s approach is a distinction between intensions and extensions, properties and sets, individual concepts and individuals; however, the distinction between the “intensional” object and its extensional correspondent is reconstructed in the a-conceptual framework of set theory: Montague’s intensions are just sets of extensions across the set of assumed possible worlds. As pointed out by Thomason in his introduction to the 1974 collection of papers of Richard Montague, “According to Montague, the syntax, semantics and pragmatics of natural languages are branches of mathematics, not of psychology.” (p. 2, Thomason, ed. 1974).

As a consequence, there is no simple connection between this semantic theory and psychology. What figures as meanings in formal semantics is nothing that can claim direct psychological reality: Our minds are finite and can handle only finite contents. There are, however, not only infinitely many possible worlds—each possible world itself is a complex of infinite information: all the information necessary to determine for all the infinitely many sentences of a language whether they are true or not. Formal semantics was never meant to provide a psychological model of meaning and semantic composition. It always aimed at capturing the logical side of language: the truth conditions for natural language sentences on the background of “worlds” taken as given, and the logical relations between sentences.

One price that the mathematical, extensional approach to meaning has to pay is fundamental: it can capture the truth conditions, more generally, the logical properties, of a sentence, but these are arguably only a derivative of the underlying conceptual level of meaning. Sentences with different meanings may have identical truth conditions. A logical approach to meaning cannot capture the differences in meaning in such cases. The most conspicuous examples are mathematical and logical truths (*two times three is six*) and analytical sentences true for just semantic reasons (*ducks are birds*) (see the contribution by de Almeida & Antal in this volume). A conceptual analysis in an intensional approach to meaning is able to capture the meanings directly, and with them the differences.

In most varieties of formal semantics, meanings are represented as expressions in an appropriate language of formal logic which is equipped with a rigid model-theoretic interpretation (other approaches formulate the truth-conditions directly). In particular, the meanings of sentences are represented by logical formulae. These formulae serve the primary purpose of formulating the truth conditions of the sentence whose meaning they represent. To give a simple (and grossly simplified) example, the meaning representation of the sentence *some spectators fainted* would be a formula

like ‘ $\exists x(\text{spectator}'(x) \text{ and } \text{faint}'(x))$ ’. What the meaning representation reflects is that there is existential quantification involved and there are two predication, ‘spectator’ and ‘faint’, applied to the same argument. Notably, the parts of the sentence that are explicitly interpreted are the functional element *some* and the predication structure of the sentence; more advanced analyses would also take care of mood, tense and aspect of the verb. Content words, however, the ordinary nouns, verbs, or adjectives, here *spectator* and *faint*, are left unanalyzed.

Formal semantics tries to account for the general rules of semantic composition, and the interplay of syntactic structure with the rules of semantic structure. For the general rules, formal semantics started out with basic logical distinctions between lexical meanings, based on logical properties that are shared by a large number of words, such as whether they denote objects, events, or properties; whether they are used for predication, and what types of arguments they predicate about. These properties constitute the “logical type” of lexical items. Semantic rules of composition essentially describe how the meanings of certain logical types of expressions combine. From this point of view, idiosyncratic differences in lexical meaning, i.e. the precise lexical meanings of individual words, do not, and should not, matter. However, for a deeper understanding of semantic composition, it turns out that one wants to know more about the expressions that combine than their logical type and their syntactic category. In Montague’s own papers, he takes care of particular words that exhibit different combinatorial properties than the “ordinary” members of this part of speech. One example is intensional verbs like *rise* in the famous construction *the temperature rises* (known as “Partee’s paradox”, see Löbner (2020) for discussion). As an intensional verb, or to be precise: in intensional use, *rise* exhibits different logical properties than verbs in extensional use, like *rise* in *the balloon rose to 30,000 m*. The intensional verb predicates about the course, or trajectory, of the temperature function, and thereby about a Montagovian “intension”, roughly the intension of the subject NP *the temperature*. By contrast, the extensional verb predicates just about a simple object, i.e. (simply speaking) about the extension of the subject *the balloon*.<sup>1</sup> Montague accounts for the logical difference between the intensional and the extensional construction by meaning postulates, not by analyzing the lexical meanings. Almost fifty years later, we are able to deal with the compositional properties of verbs like *rise* on the basis of a decomposition of their meaning (see the contribution by Gamerschlag & Petersen in this volume). The decomposition explains how the verb meaning interacts with its arguments in different constructions, intensional and extensional, resulting in sense variation of the verb. The analysis of the lexical meaning of the verbs predicts the compositional behavior of this (and similar) verbs.

Natural language semantics, ultimately, needs to provide theories and analysis of lexical meaning, not only of general rules of semantic composition. This is the more so as formal semantics has long since taken a course of constant differentiation, turning to more and more detailed problems, ever closer to the analysis of phenomena that hold only for a small number of words, if not sometimes for a single word. Ideally,

---

<sup>1</sup> Montague’s formal solution is in terms of more complex logical types, but it is logically equivalent to the simplified picture given here.

a theory of semantic composition would start out from decomposition—a description of the structure and content of lexical meanings—and proceed to describe how they combine in a given syntactic construction. The starting point of this endeavor, the analysis of lexical meanings, is, however, an arduous enterprise: there are so many words; each of them potentially with different senses, resulting in hundreds of thousands of lexical meanings in a language like English. Thus, it makes sense to first start out with very coarse semantic distinctions such as the logical type (like ‘n-ary predicate expression’, ‘quantifier’, ‘logical connective’, and so on). Beyond that, most developments of formal semantics have investigated lexical meanings of content words only to a very limited extent.

There are a few exceptional forays by formal semanticists into the realm of lexical meanings, notably Dowty’s decomposition of different types of verb (Dowty 1979) which became widely accepted. Otherwise, lexical semantics remained a stepchild of formal semantics; the discipline never came up with a general framework for decomposition. A later proposal for a more general approach to the decomposition of lexical meaning was presented in Pustejovsky’s (1995) theory of the “Generative Lexicon”. It was extended substantially in many follow-up case studies. The theory proposes a general structure of lexical meanings in terms of four qualia that capture focal properties of the potential referents including form, purpose, origin, along with argument structure and event structure for verbs. The theory models lexical meanings not only of verbs, but also of nouns. The structure of the lexical entries can be considered some variant of frame; Pustejovsky’s lexical meanings are, however, considerably more restricted than general Barsalou frames. Pustejovsky’s theory of the lexicon is an influential and very important development in linguistic semantics. For many phenomena, it is able to model semantic composition in a much more detailed and differentiated way. This is possible because there is so much more information on the lexical meanings available. Pustejovsky convincingly demonstrated that any detailed theory of semantic composition ultimately needs to be based on decomposition if one wants to better understand how the meanings of the components of a complex expression combine.

However, even with decompositional elements and an apparatus like Pustejovsky’s Generative Lexicon, mainstream formal semantics never developed into a psychological (or cognitively oriented) theory of meaning. With the growing influence of cognitive psychology, attempts at connecting linguistic semantics to the facts and theory of cognition have been gaining considerable momentum (see, e.g., Murphy 2002, Chap. 11). This development is in the interest of both semantics and cognitive psychology. If one assumes that linguistic meanings correspond to concepts stored in the cognitive system, then semantic analysis can yield insights into the architecture and mechanisms of the cognitive system, and the empirical investigation of the latter can provide stronger, and different, criteria for adequate semantic analysis.

A theory of linguistic meanings as structures stored or formed in the cognitive system, requires a theory of representations of meanings and concepts in general. One of the goals of the Düsseldorf CRC 991 was to develop a frame theory as a generally applicable theory of representations. The origin and point of departure is Barsalou’s theory of frames which he claimed are a candidate for the general format of cognitive

representations (Barsalou 1992a, b). The CRC research has applied Barsalou's frame hypothesis to language, for the modeling of linguistic representations in semantics, syntax, morphology and phonology (see Löbner 2014 for a general discussion of the consequences of the frame hypothesis for the understanding of language). Other scholars applied the approach in their philosophical and psychological research.

Many contributions in this volume take a position with respect to the relationship between meaning and concepts for the issue of decomposition. There's the extreme position argued for by de Almeida & Antal, who argue against decomposition. In their model of natural language semantics, lexical meanings are stored units not to be decomposed, i.e. atoms in the semantic system.

While formal semanticists mostly have practiced lexical atomism by assuming that lexical meanings are just given as they are, they would not argue against decomposition if necessary and feasible. This practice is to be observed in the three formal semantics contributions by Morzicky, Krifka, and Liefke, whose concern is not so much with lexical meanings and their interaction but with the interpretation of certain constructions. Robin Cooper's contribution is in a similar vein as far as lexical decomposition is concerned. He develops a remarkable theory of connecting semantics and cognition and accounting for semantic phenomena with complex cognitive structures, but these structures still contain unanalyzed lexical meanings. At the opposite end of the scale, there are frame-based semantic analyses (Andreou & Petitjean, Balogh & Osswald, Gamerschlag & Petersen, Löbner). These contributions propose frame-based decompositional structures as the basis of modelling semantic composition for a variety of phenomena. Berio applies the frame approach to her discussion of the meaning of color terms.

## 2 Cognitive Structures in Philosophy

In the introductory part on natural language semantics, we sketched Montague's semantics and mentioned Gottlob Frege, one of the founding fathers of philosophy of language and of linguistic semantics in general. Indeed, Frege's notions of *Sinn* (sense) and *Bedeutung* (reference) are what Montague intends to capture with his notions of intension and extension, using Rudolf Carnap's development of possible worlds in, for example, his *Meaning and Necessity* (1947). Moreover, Frege already formulated the central semantic principle of compositionality which we find in Montague and in Alfred Tarski's work on the truth predicate for formal languages (1936). It was also taken up by Donald Davidson (1967) to introduce truth-functional semantics for natural languages: the meaning (truth-conditions for sentences) of a complex expression is a function of the meaning of its parts and the way these parts are put together in the expression. In fact, most of those who built the foundations of formal semantics were not linguistic semanticists, but philosophers, such as Frege, Carnap, Tarski, Davidson, Montague, Lewis or Cresswell, to mention only a few. Barbara Partee and Robin Cooper are among the early protagonists with a genuine linguistic background; Robin Cooper is one of the contributors to this volume. Formal

semantics with its background of analytic philosophy and logic was tremendously important for linguistics because it helped to establish semantics as one of its central disciplines. However, the extensional turn—the replacement of Frege’s *Sinn* by the mathematical notion of intension severed the discipline from a conceptual, that is psychological point of view. This made it difficult to connect mainstream semantics to the developments in cognitive science.

The emergence of modern cognitive science is the arrival of computational models within cognitive psychology, models that are inspired by logic, philosophy, linguistics, and artificial intelligence, and required intensive collaboration between logicians, philosophers, linguists, psychologists, and computer scientists. One of these models, of particular influence for many contributions in this volume, is Barsalou’s frame model; it “borrows heavily from previous frame theories, although its collection of representational components is somewhat unique”.<sup>2</sup> Cognitive structures belong to cognitive science in the sense described above where cognitive science is meant to improve the understanding of human cognitive skills, like categorizing, learning, reasoning and planning, and by developing better and better models of these skills, models that—if they are not directly implemented—clearly could contribute to implementation if existing limitations were removed. Modeling concepts and other cognitive structures is a core enterprise.

Theories of concepts have been central in philosophy for as long as it is practiced as a discipline. One of the most important, if outdated theories is the one found in Locke and Hume, but related to a tradition going back to Aristotle where concepts are identified with images or (pictorial) representations. Another classical view—recently defended again by Peacocke (1992)—takes the necessary and sufficient conditions for the application of a concept to an instance as identity criterion for a concept. A modern alternative to such classical theories is the so-called theory theory of concepts (Gopnik and Meltzoff 1997) in which an analogy is made to the meaning of theoretical terms in scientific theories and in which the content of concepts is given by the theories in which they figure. The exemplar theory of concepts (Brooks 1978) starts from classification learning and defines the extension of the concept as the class of objects which are sufficiently similar to typical exemplars. Rosch (1978) develops a prototype theory of concepts in which objects fall under a concept if they match with a prototype to a certain degree. This view can be related to the family resemblance theory of Wittgenstein. The approach most elaborate on representation is Barsalou’s (1992a, b, 1999) frame theory of categorization. For the Düsseldorf CRC 991, Barsalou’s frame theory is the central candidate for a theory of cognitive conceptual representations and means of categorization.

The success of cognitive science research also means that improvements in cognitive modelling can lead to new insights within the disciplines that inspired the first versions of the models. In the case of logic and philosophy, the contribution to cognitive science ranges over a number of areas. The development of formalizations of logic for the mathematical study of logic has led to precise versions of notions such

---

<sup>2</sup>Barsalou (1992a, p. 21). In Barsalou (1992b, p. 158), he mentions various sources from linguistics, artificial intelligence and logic.

as proposition, proof, entailment, contradiction, tautology, validity, completeness, and others that can be used as first models of human inferential and representational skills, to be tested against empirical data.

Alvin Goldman's is a different kind of contribution from philosophy to cognitive science. His theory of human action (1970) turns out to provide a novel general, very far-reaching, model for the cognitive theory of categorization. According to Goldman, human action very often constitutes simultaneous action at many levels. His theory was presented as a contribution to ontology, but in reply to his critics he later stated that it is in fact a psychological theory of categorization (see Löbner's chapter in this volume).

There is an increasing number of philosophers of mind and of language who are themselves cognitive science researchers (or at least follow cognitive science research closely), among them Alvin Goldman (with more recent work), Peter Hanks, Thomas Metzinger, Friederike Moltmann, Albert Newen, Elisabeth Pacherie, Josef Perner, François Recanati, Gottfried Vosgerau and Markus Werning. While this research may be directed at new results or new arguments within ongoing philosophical discussions, it is nonetheless straight cognitive science, even if the questions addressed do not come directly from a psychological cognitive science agenda.

### 3 Cognitive Structures in Psychology

The ability to form conceptual representations has been a core research interest in psychology since the cognitive revolution almost half a century ago. Much of the theoretical and empirical work in cognitive psychology is, and has been, influenced by parallel research lines in philosophy and natural language semantics, some of which are mentioned above. One example is the classic feature list model in cognitive psychology that was developed by Glas & Holyoak (1975) and Hampton (1979). They proposed that each category representation is a list of features, that is, a list of independent representational components forming a single level of analysis, whose sum represents the category. Feature lists treat attributes and values as the same kind and do not specify relations between features. By contrast, as outlined above, frame theory according to Barsalou and others (Barsalou 1992a, 2005) is supposed to be an alternative to flat feature list representations, but also to other theories prominent in the research literature such as prototype theory and exemplar theory. The frame approach holds that concepts can be represented in attribute-value structures. Each attribute can be connected to a cluster of more specific attributes, and certain attributes can also constrain the range of other attributes putting the concepts into dynamic connection and relation. One implication is that the activation of a perceptual property of a concept in frame format may automatically lead to the representation of a whole conceptual system, which allows a structured description of knowledge (Barsalou 2005).

The feature- or attribute-list framework has been hypothesized to be species-general. Referring to the work of Sutherland and Mackintosh (Mackintosh 1965;

Sutherland & Mackintosh 1971), Barsalou already proposed in 1992 (Barsalou 1992a) that not only humans, but non-human animals, too, use attribute-value sets to conceptually represent their world, and he more recently made the claim of a continuity of the conceptual system across species more specific (Barsalou 2005). For example, in a rat version of the set shifting task (Birrell & Brown 2000), animals had to choose between two different bowls where one contained a food reward, and the other did not. The bowls differed in three attribute values: odors, mediums that filled the bowl, and surface textures. One of these attributes cued which of the two bowls contained the reward. Once rats learned to identify the reward-predicting cues, the cue-reward contingencies were shifted. Results showed that learning a novel discrimination was faster in so-called intradimensional shifts when the discrimination was based on the previously relevant perceptual dimension (e.g. odor–odor cue reversals: oregano to cinnamon) compared with a condition when attention had to be shifted to the previously irrelevant dimension in so-called extradimensional shifts (e.g., odor–filling reversals: oregano to sand). The shift-costs, i.e., the post-reversal reacquisition rate, should be identical after intra- and extradimensional shifts if the cue was represented as a feature list. However, this was not the case: the animals were slower to reach pre-shift performance after an extra- compared to an intradimensional shift. This observation is difficult to explain with the hypothesis of isolated feature list representations. A better way to understand these phenomena is that the stimulus is represented by each of its attributes and attribute values, e.g. “odor” with the values oregano or cinnamon. A shift between the values of the same attribute should be easier than a shift between different attributes. The chapter by David Tait, Verity Brown and colleagues in this volume stands in the tradition of this research, and investigates the neural mechanism underlying reversal learning in rats.

It has recently even been argued that frame theory can be extended to understand conceptual representations of animals in the social domain. For example, Gil-da-Costa et al. (2004) studied macaques, and investigated the cognitive and neural representation of social calls emitted by conspecifics. They found that the calls conveyed information about the caller and its socioecological context. There were two types of calls: the first was named *coos* and was associated with positive social context, such as friendly approach behavior. The second type was termed *screams*, which are usually emitted in threatening situations, such as an attack by a conspecific. By using Positron-Emission Tomography, it was found that these conspecific vocalizations elicited activity in neural networks that strongly correspond to the network shown to support the representation of conspecifics and affective information in humans. The chapter by Kalenscher and colleagues in this volume expands on this finding, and argues that conspecifics’ calls in rats evoke multi-level representations by carrying acoustic and motivational value; they can, thus, structure rat social interaction.

These examples show that cognitive and comparative research can yield insights into a universal representation system of cognition that applies across species and domains. Hence, bringing together theoretical and empirical work from philosophy, natural language semantics and cognitive comparative psychology bears synergies that either discipline alone could not achieve.

## 4 Summaries

### 4.1 *Part I Pushing the Boundaries of Formal Semantics*

This part consists of contributions by formal semanticist, which—in one or the other way—undertake to push the boundaries of present formal semantic theory. They push the boundaries in different respects and in different directions. There is the general challenge to the truth-conditional model-theoretic approach that formal semantics is taking (invariably from its early beginnings until today), that it is intrinsically noncognitive, assuming essentially an idealized omniscient epistemic perspective on truth and truth-conditions. In an early paper on the nature of the Montagovian approach, Barbara Partee posed the question “Semantics—mathematics or psychology?”, where she observes that Montague semantics is a mathematical method of doing semantics and modeling meaning; however, she points out, attitude reports seem to require a psychological perspective on their semantic analysis (Partee 1979). We reencounter an aspect of the problem in Liefke’s attempt to include the existence of subjective cognitive systems into a wider framework of formal semantic analysis of belief sentences. Counting of various logical types of things has been a challenge to logical analysis and the ontological design of the framework of possible-worlds semantics (cf. Krifka’s classical 1990 paper “Four thousand ships passed through the lock: Object-induced measure functions on events”). In Krifka’s contribution to this volume, we will tackle with temporary configurations. A different challenge is the assumption of the homomorphism of morphosyntax and semantic composition. It was a central topic since Montague’s first treatment of quantification in 1973 which proposed a formal solution to the seeming incongruence of syntactic and semantic structure in the case of nominal quantification. Certain types of seemingly displaced adjectives remain a challenge to date (cf. the paper by Morzicky in this volume).

Kristina Liefke’s chapter “[A Compositional Pluralist Semantics for Extensional and Attitude Verbs](#)” proposes a new account of linguistic content that reconciles content-pluralism with compositionality. This is achieved by integrating truth-conditional content and attitude report content into a single notion of content. A parametrized version of this notion (with parameters for agents, times, and information states) serves as input to the compositional semantic machinery. By supplying different parameter-values to the parameterized contents of their complements, different verbs select for different components of the complement’s integrated content. The resulting account explains the different substitution properties of extensional and attitude constructions and captures the role of agents’ epistemic perspective in the determination of attitude content. The account improves upon other accounts of truth-conditional and attitude content (esp. two-dimensional semantics) by interpreting different occurrences of an expression—in extensional and in attitude embeddings—as objects of the same semantic type, and by explaining the substitution-resistance of attitudinal embeddings of extensional constructions.

**Manfred Krifka**, in his contribution “[Counting Possible Configurations](#)” deals with entities such as outfits: these consist of a configuration of pieces of clothing; they come into existence when actually combined, cease to exist when not worn, and may or may not come into existence again. To count how many outfits one has is a challenge to formal semantics, as it is often assumed that a requirement for counting objects is that they do not overlap. This condition is violated in cases such as outfits. The article develops an analysis of such configurational entities as individual concepts. It investigates the interaction of noun phrases based on such nouns with modal operators and in collective and cumulative interpretations. The general direction of this paper points towards a theoretical framework in which the objects referred to in language, and consequently, the objects of our cognition, should be seen as individual concepts. The notion of an object contains the ability to identify the same object over different indices, and this is precisely achieved by individual concepts. Some objects are temporally convex in the sense that they have a continuous existence from an initial time to a final time (such as shirts and pants), others have a more spotted existence (such as outfits).

**Marcin Morzicky**’s concern is with cases of adjective constructions that appear to provide notorious problems to the assumption of a match between grammatical and semantic structure. In his paper “[Structure and Ontology in Nonlocal Readings of Adjectives](#)”, he refers to them as adjectives with “nonlocal” readings, i.e. readings in which the adjective (for example *occasional* or *average*) appears to make the contribution of an adverb. Morzicky points out that the phenomenon is more general than usually assumed. There are two options, he argues, to deal with this kind of phenomenon: to invest into a richer and maybe cognitively more ambitious ontology and to invest in more involved composition rules. As to the intuition that these nonlocal adjective readings are a grammatical oddity, Morzicky concludes: “These adjectives are indeed odd, but in a precise and interesting sense. They are odd in the way that platypuses and lungfish are odd: they are—perhaps metaphorically, or perhaps more than metaphorically—transitional forms in an evolutionary progression, unusual because they combine features of two distinct categories that we normally regard as mutually exclusive.”

## 4.2 Part II Concept Theory

The papers in this section provide more general accounts of how one can approach the nature of concepts from a formal point of view. They deal with very essential questions: Should the meaning of lexical items be approached by means of decomposition/internal analysis or rather be treated as atomic/opaque? How is the concept space structured and what makes a “natural” concept? How is categorization related to perception and which system of types does one have to assume in this regard? What’s the impact of language on concepts? The contributions in this section show that these questions—in spite of their classic nature—are at the very heart of present-day research on concepts, meaning and representation.

In their contribution “[How Can Semantics Avoid the Troubles with the Analytic/Synthetic Distinction?](#)” **Roberto G. de Almeida** and **Caitlyn Antal** present a criticism of semantic theories that differentiate between analytic and synthetic features, a distinction originally grounded in the philosophical opposition between statements that are logically true and those whose truth depends on additional world/contextual knowledge (Kant 1781; Carnap 1956). In favor of their opinion, de Almeida and Antal discuss potential problems of the lexical decomposition account of causative verbs and the type-coercion analysis of semantic mismatches between verb and argument meaning. As an alternative to these accounts, the authors sketch analyses based on the assumption that concepts invariably contribute all of their contents and do not involve a characterization by features (“concept atomism”). They show how some of the regularities found with causatives as well as type-coercion can be analyzed in terms of inferences/meaning postulates triggered by the meaning of lexical items.

**Leda Berio** discusses the way conceptual representations can be conceived of as being determined by language in her chapter “[Linguistic Relativity and Flexibility of Mental Representations: Color Terms in a Frame Based Analysis](#)”. She argues that Whorfianism/language relativity on the one hand and universalism on the other hand are extreme oppositions one of which needs not be necessarily assumed given more recent developments which offer a more differentiated, less radical picture of the interrelation between language and concept formation. As a format of mental representation and a device for mediating between linguistic and perceptual information in concepts, Berio proposes frames in the sense of Barsalou (1992a, b) and Löbner (2015). She shows that frame representations exhibit a high degree of flexibility which allows for the representation of the interaction between linguistic and perceptual information necessary to capture the results of experiments related to the relativity/universalism debate, in particular those dealing with color labeling.

Starting from the major division into conventional and conversational implicatures and following subtypologies such as the differentiation between various kinds of scalar implicatures which have developed as some kind of mainstream after the original definition of the term by Grice (1975), **Igor Douven** investigates the conceptual properties of implicatures in his paper “[Implicatures and Naturalness](#)”. In particular, Douven is interested in the question whether implicatures should be regarded as natural concepts having a reality independent of what he refers to as “linguistic intuitions”. The author proposes to deal with that question in terms of Gärdenfors’ theory of conceptual spaces (Gärdenfors 2000) and to check whether different kinds of implicatures satisfy Gärdenfors’ “Criterion P” that a natural concept is a convex region of a conceptual space. Based on data from a self-conducted study, Douven constructs a conceptual space for different types of implicatures and argues that the distribution of items in the implicature space suggests a characterization of implicatures as natural concepts.

In his chapter “[Perception, Types and Frames](#)”, **Robin Cooper** offers an approach to perception and categorization formulated within his framework of Type Theory with Records (TTR, Cooper 2012). He claims that perception is determined by the

way we classify entities (i.e. objects and events) according to this framework. Characteristically, TTR goes beyond the traditional binary distinction between entities and truth values put forward by Montague (1974) in building on a more elaborate system of types following the type theory of Martin-Löf (1984). Thus, TTR also assumes basic types for physical objects and events. Cooper gives an introduction to the essentials of TTR with special reference to the conception of “record types” and their instantiation by particular records both of which play a central role within this theory. Moreover, Cooper discusses how his model is related to Fillmore frames and to cognitive frames in the sense of Barsalou (1992a, b) and their formal adoption by Löbner (2014, 2015), Kallmeyer & Osswald (2013) and Kallmeyer et al. (2017) among others.

### **4.3 Part III Conceptualizing Eventualities**

Eventualities are temporal entities, usually understood as comprising events and states both of which have a temporal structure and a location in time. According to Guarino (1997) eventualities can be characterized as ‘occurrents’ which differ ontologically from ‘continuants’ defined as objects lacking both temporal location as well as temporal parts while characteristically exhibiting ‘mereo-topo-morphological properties’. Both types of entities are closely related to each other such that “‘occurrents are ‘generated’ by continuants, according to the ways they behave in time” (Guarino 1997: 7). The papers in this section deal with different aspects of eventualities and the way they are conceptualized. Since events are referred to characteristically, but not exclusively, by verbs, all contributions are concerned with phenomena related to verbs such as deverbal nominalizations, verbal aspect, verbal particles and stative readings of dynamic verbs. The last chapter proposes a cognitive structure for representing action, and thereby the meaning of action verbs: the model of so-called cascades. It is based on Goldman’s multi-level account of human action that assumes that action more often than never is to be categorized simultaneously at different levels.

In their paper “[An XMG Account of Multiplicity of Meaning in Derivation](#)” **Marios Andreou** and **Simon Petitjean** propose an account of the various readings exhibited by English deverbal nouns resulting from *-al*-suffixation. Based on a corpus study, the authors show that apart from an event and result reading *-al* derivatives can display also readings of a non-eventive nature which refer to a variety of participants involved in the event denoted by the base verb. The different readings which are available (or excluded) for a specific verbal base are captured by type constraints which single out particular components in a frame representation of the base verb as referents of the nominalization. One merit of this approach is the reduction of over-generation, a problem characteristic of monosemous accounts of derivation which assume a general underspecified meaning for an affix. In the final part of their paper, Andreou and Petitjean offer a formalization of their analysis by modelling it using Extensible Metagrammar (XMG, Crabbé et al. [2013](#)).

**Martín Fuchs, Ashwini Deo and María Mercedes Piñango** discuss the way nonlinguistic constraints determine the use of aspect markers in their contribution “[Operationalizing the Role of Context in Language Variation: The Role of Perspective Alignment in the Spanish Imperfective Domain](#). ” The authors start out from the results of a study on the relevance of the context on the availability of the simple present as a marker of progressive meaning as opposed to the context-independent accessibility of the present progressive marker in three different varieties of Spanish. Fuchs et al. propose an account which builds on a process they call ‘perspective alignment’. Perspective alignment aims at bringing the hearer’s perspective closer to the speaker’s perspective. According to the authors, this process can be considered as mediating between the opposite principles of linguistic economy and linguistic expressiveness. In particular, the progressive interpretation of the simple present in Spanish is only available if speaker and hearer both have perceptual access to the event denoted by the verb which ensures the speaker-hearer perspective alignment in a non-linguistic way.

In “[A Frame-Based Analysis of Verbal Particles in Hungarian](#)” **Katalin Balogh** and **Rainer Osswald** provide a formal approach to the semantic contribution of the Hungarian particles *meg-*, *le-*, *el-*, and *fel-* and the way they combine compositionally with their respective verbal base. In their account, they apply a formalization of Role and Reference Grammar (Van Valin & LaPolla 1997) on the one hand and a decompositional frame semantics as a device for combining lexical decomposition with a frame representational format on the other hand. The explicit formalization of the semantic interaction between verbal base and particle sets their approach apart from previous approaches to Hungarian particles which do not elaborate formally on the semantic and syntactic representation of the base verb and the particle and the way they are combined in a compositional semantics. A further aspect addressed by the authors is the syntactic distribution of verbal particles and resultative phrases and how these patterns can be analyzed compositionally by means of frame semantics.

In their paper “[On the Fictive Reading of German \*Steigen\* ‘Climb, Rise’: A Frame Account](#)”, **Thomas Gamerschlag** and **Wiebke Petersen** deal with the stative use of verbs of motion frequently referred to as ‘fictive motion’ (Talmy 2000). The authors present a case study of the fictive motion reading of the German movement verb *steigen* ‘climb, rise’ and show how it can be analyzed by contrasting it to the dynamic readings of the verb within a frame account. In particular, they argue that both the fictive motion reading as well as the so-called ‘intensional’ reading of *steigen* derive from the non-figurative directional reading of the verb since all of these readings obligatorily exhibit a value change restricted to a positive difference. In Gamerschlag and Petersen’s frame account, the intensional and the fictional uses result from different operations on the frame representation of the directional use (replacement of the POSITION-attribute in the former case vs. deactivation of the dynamic frame components and accommodation of the meaning of the subject in the latter case).

**Sebastian Löbner**’s contribution “[Cascades, Goldman’s Level-Generation, Multilevel Categorization of Action, and Multilevel Verb Semantics](#)” proposes a novel theory of the categorization of acts and applies it to the semantics of action

verbs, with fundamental consequences for semantic theory and beyond. The theory is based on Goldman's (1970) multilevel theory of action which is taken here as a theory of categorization. Goldman's central notion is *level-generation*: acts of a type may under circumstances generate acts of other, more abstract, types. The acts form a hierarchical structure that Goldman calls an *act-tree*. Level-generation results in a conceptual relation called *c-constitution* here, i.e. constitution under the given circumstances. Löbner introduces the more general term *cascade* for act-trees. In the second part, multilevel cascade-structure categorization is conflated with a cognitive semantics that models meanings with Barsalou frames. A multilevel analysis of the concept of writing is discussed in depth and detail in order to illustrate the potential and the consequences of a cascade approach to verb semantics. It is shown that the concept of c-constitution can be generalized as to cover the roles of persons and objects across levels in a cascade. The generalization suggests that multilevel categorization may be a very general and fundamental phenomenon in the psychology of categorization.

#### **4.4 Part IV Prototypes and Probabilities**

It is a well-known phenomenon that human cognition is able to recognize less-typical specimens as belonging to a particular category although they differ more or less drastically from the perfect representatives of this category (Rosch & Mervis 1975; Rosch 1978). From a theoretical point of view, the challenge in this regard is to capture the relevant cognitive factors underlying the process of categorization and in particular to provide suitable mechanisms able to deal with the non-representative instances of a category. The contributions in this section offer approaches to the categorization and comparison of individuals which deal with the question how the underlying concepts are structured. Characteristically, all of these accounts assume representations of a much more elaborate structure than the feature lists of early prototype theory.

**Corina Strößner, Annika Schuster and Gerhard Schurz** discuss the effect of modification on prototype compositionality in their paper “[Modification and Default Inheritance](#)”. Starting from the observation that modification characteristically leads to a decrease of how likely typicality statements are rated, the authors propose an account of prototype composition in adjective-noun combinations as a representative pattern of modification. Their analysis is based on an extension of the selective modification model by Smith et al. (1988). In particular, Strößner et al. add the expressivity of Barsalou frames (Barsalou 1992a, b) which allows for capturing cross-attributional constraints, i.e. co-variation of different attributes of an entity such as the indication of a sour TASTE of an apple by its green COLOR. The formal approach is complemented by an exploratory study in which participants rated the typicality and likelihood of properties of modified and unmodified nouns as well as the typicality and likelihood of particular modifiers of a given noun.

**Samuel Taylor** and **Peter Sutton** present a frame approach to Bayesian models of categorization in their article “[A Frame-Theoretic Model of Bayesian Category Learning](#)”. They claim that frame representations are advantageous over unstructured feature list representations which are commonly applied in Bayesian models. In particular, Taylor and Sutton argue that it is a shortcoming of the use of feature list representations that they usually depend on supervised training data for assigning weights to features. As an alternative, they introduce frame representations for mediating between sensory input and behavioral output and show that the recursive structure of frames can be exploited in a way which allows for the weighting of attribute values in an unsupervised process of categorization. By analyzing a simple example of animal categorization, the authors demonstrate that attribute values can be weighted in terms of their appearance in the frame: features belonging to attributes closer to the central node of a frame are more important and are assigned more weight than features of attributes located more distant from the central node of a frame.

In their contribution “[Extremes are Typical. A Game Theoretical Derivation](#)”, **Robert van Rooij** and **Thomas Brochhagen** challenge the hypothesis that a prototype understood as a typical specimen of a category is also a central member of that category. By contrast, the authors claim that rather stereotypes which are defined as extreme exemplars constitute the typical instances of a category. Consequently, although they follow Gärdenfors’ (2000) idea that basic categories are always convex sets, they oppose his assumption that prototypes are at the center of a convex set. By discussing color and taste space as basic examples of Gärdenfors’ theory of conceptual spaces, Rooij and Brochhagen argue that typical representatives of color and taste are at the edges of the respective spaces and “as far away from each other as possible”. In line with their assumption, they propose a game theoretic analysis in which both convexity of meaning as well as stereotypes are accounted for as resulting from principles of rational language use.

In deciding whether an entity belongs to a particular category, similarity of objects plays a central role. In their paper “[Grading Similarity](#)” **Carla Umbach** and **Helmar Gust** present an analysis of the German/English similarity expressions *ähnlich/similar*, *so/such*, and *gleich/same* with a particular focus on the explanation of gradability asymmetries (*ähnlich/similar* are gradable expressions in contrast to *so/such* and *gleich/same*). The authors propose an approach to similarity in which the three different expressions of similarity in German and English are treated by means of a similarity relation  $SIM(x, y, \mathcal{F})$  with  $\mathcal{F}$  being defined as a quadruple comprising the domain of entities, an attribute space, a measure function and a set of classifiers. Umbach and Gust argue that the use of the similarity expressions under discussion can be analyzed by considering in particular the set of classifiers and the different dimensions of comparison which are associated with a specific attribute space. Their account of the gradability of *ähnlich/similar* is motivated by ideas originally put forward in Klein (1980).

## 4.5 Part V Cognition and Psychology

This part addresses the question of cognitive structures from an empirical perspective that applies not only to human cognition, but also to the cognition of rats. Both contributions on rat psychology address basic questions of cognitive structures concerned with cognitive mechanisms that play a role in reinforcement learning. One of the “human” contributions concerns the interaction of language processing with the cognitive motor system. The study differentiates and corroborates the findings on the embodiment of semantic knowledge first reported in Pulvermüller (2005) and in many later studies. The other addresses the radical question whether cognitive representations should be assumed to exist at all.

In their paper “[Escitalopram Restores Reversal Learning Impairments in Rats with Lesions of Orbital Frontal Cortex](#)”, **David Tait, Ellen Bowman, Silke Miller, Mary Dovlatyan, Connie Sanchez and Verity Brown** investigate the neural underpinnings and the malleability of cognitive structures. Cognitive structures can be defined as mental models, and they improve the efficiency of information processing by providing a situational framework within which there are parameters governing the nature and timing of information. Tait, Brown and colleagues study cognitive structures by training rats in a reversal learning task where previously acquired stimulus-response contingencies are reversed, and subsequently reverted to the original contingency. Lesions of the rats’ orbitofrontal cortex resulted in poorer reversal performance. For example, they showed higher perseveration errors (the rats continued to choose the previously rewarded, now unrewarded cue after a reversal) and took longer to acquire the novel stimulus-response contingencies after a reversal. This impairment in reversal performance was restored to normal performance by administration of escitalopram, an antidepressant drug that increases the synaptic transmission of the neurotransmitter serotonin. In addition, the orbitofrontal cortex lesions resulted in an increase of neuronal activity markers in prefrontal regions, which were even more amplified by escitalopram administration. These results suggest that cognitive structures, enabling learning by representing the world as a cognitive map, involve orbito- and prefrontal brain structures, and can be modulated by serotonergic action.

The contribution by **Tobias Kalenscher, Lisa-Maria Schönfeld, Sebastian Löbner, Markus Wöhr, Mireille van Berkel, Maurice-Philipp Zech and Marijn van Wingerden** deals with rats psychology, too. In their paper “[Rat Ultrasonic Vocalizations as Social Reinforcers—Implications for a Multilevel Model of the Cognitive Representation of Action and Rats’ Social World](#)”, the experimental focus is on prosocial behavior; the second part offers a cognitive modelling of reinforcement learning as cascade formation. The empirical research investigated the role of certain ultrasonic vocalizations (USV) which rats produce at frequencies of either 50 or 22 kHz. The chapter presents evidence supporting the hypothesis that USVs act as social reinforcers. In line with the social reinforcement hypothesis (Hernandez-Lallement et al. 2017), it is shown that rats preferred T-maze compartments associated with 50-kHz USV playback over compartments associated with non-ultrasonic control stimuli. This observation fuels the hypothesis that USVs might orchestrate

and structure social interaction between rats. From the point of view of cascade theory (cf. the contribution by Löbner in this volume), ultrasonic vocalizations with a social “meaning” are assumed to be represented in the rat’s brain as two-level cascades with a lower, physical, level of vocalizing and a higher, social, level of signaling. The main application of cascade theory is to the modeling of reinforcement learning, considering it as the formation of a cascade that invests a particular behavior with the aspect of making oneself have a rewarding or aversive experience. This model of learning would explain the acquisition of practical knowledge-how as the result of a basic brain mechanism of cascade formation. This is important in the given context because the same cognitive learning mechanism is very plausibly to be observed with human subjects, too, in their acquisition of the daily knowledge-how. Thus, it appears, cascade formation is a basic brain mechanism across species.

**Jan Sieksmeyer, Anne Klepp, Valentina Niccolai, Jacqueline Metzlaff, Alfons Schnitzler, and Katja Biermann-Ruben**’s contribution “[Influence of Manner Adverbs on Action Verb Processing](#)” aims to investigate motor cortical involvement in the processing of hand- and foot-related action verbs combined with manner adverbs, applying behavioral methods and EEG recordings. The study provides an indication that manner adverbs influence motor behavior while corroborating the already existing data concerning the interaction between action verb processing and motor output. These findings are in line with assumptions made by embodied cognition theories proposing an essential role of sensorimotor areas in the processing and storage of action concepts inherent in action-related language. The adverbial modulation of motor behavior might reflect a certain variation of motor involvement in language processing. This involvement could be susceptible to grammatical constructions modifying the action component of action verbs. Yet, effects of the verb material in a closely matched verb set and influences of timing have to be taken into account.

In his paper “[When Mechanical Computations Explain Better](#)” **Silvano Zipoli Caiani** discusses the position of radical enactivism (e.g. Hutto and Myin [2012](#)) whose supporters argue that the representational-computational paradigm does not add explanatory power over and above the physical description of a cognitive system, and therefore should be abandoned. Zipoli Caiani defends the representational-computational paradigm in a careful study of the phenomenon of *optic ataxia*, a disorder characterized by difficulties in executing visually-guided reaching tasks, although ataxic patients do not exhibit any specific disease of the muscular apparatus. He demonstrates that the assumption of the dual stream model of vision—and hence a computational brain mechanism—explains phenomena that the radical enactivism paradigm is unable to account for.

## References

- Barsalou, L. W. (1992a). Frames, concepts, and conceptual fields. In A. Lehrer & E. F. Kittay (Eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization* (pp. 21–74). Hillsdale NJ: Lawrence Erlbaum.
- Barsalou, L. W. (1992b). *Cognitive psychology. An overview for cognitive scientists*. Hillsdale NJ: Lawrence Erlbaum.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Barsalou, L. W. (2005). Continuity of the conceptual system across species. *Trends in Cognitive Sciences*, 9, 309–311.
- Birrell, J. M., & Brown, V. J. (2000). Medial frontal cortex mediates perceptual attentional set shifting in the rat. *Journal of Neuroscience*, 20, 4320–4324.
- Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale NJ: Lawrence Erlbaum Associates.
- Carnap, R. (1956 [1947]). *Meaning and necessity*. Chicago IL: Chicago University Press.
- Cooper, R. (2012). Type theory and semantics in flux. In R. Kempson, N. Asher, & T. Fernando (Eds.), *Handbook of the philosophy of science. Philosophy of Linguistics* (Vol. 14, pp. 271–323). Oxford, Amsterdam: Elsevier.
- Crabbé, B., Duchier, D., Gardent, C., Le Roux, J., & Parmentier, Y. (2013). XMG: eXtensible MetaGrammar. *Computational Linguistics*, 39, 1–66.
- Davidson, D. (1967). Truth and meaning. *Synthese*, 17, 304–323.
- Dowty, D. R. (1979). *Word meaning and Montague Grammar*. Dordrecht: Reidel.
- Fillmore, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, 280, 20–32.
- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* C: 25–50. English translation (1948): Sense and reference. *The Philosophical Review*, 57, 209–230.
- Gärdenfors, P. (2000). *Conceptual spaces. The geometry of thought*. Cambridge MA: MIT Press.
- Gil-da-Costa, R., Braun, A., Lopes, M., Hauser, M. D., Carson, R. E., Herscovitch, P., & Martin, A. (2004). Toward an evolutionary perspective on conceptual representation: species-specific calls activate visual and affective processing systems in the macaque. In *Proceedings of the National Academy of Science of the U.S.A.*, 101, 17516–17521.
- Glas, A. L., & Holyoak, K. J. (1975). Alternate conception of semantic memory. *Cognition*, 3, 313–339.
- Goldman, A. I. (1970). *Theory of human action*. Princeton NJ: Princeton University Press.
- Gopnik, A., & Meltzoff, A. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.
- Grice, H. P. (1975). Logic and conversation. Reprinted in H. P. Grice (Ed.), *Studies in the way of words* (pp. 22–40). Cambridge MA: Harvard University Press (1989).
- Guarino, N. (1997). Some organizing principles for a unified top-level ontology. In *AAAI Spring Symposium on Ontological Engineering* (pp. 55–63). Menlo Park CA: AAAI Press.
- Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 18, 441–461.
- Hernandez-Lallement, J., van Wingerden, M., Schäble, S., & Kalenscher, T. (2016). A social reinforcement learning hypothesis of mutual reward preferences in rats. *Current Topics in Behavioral Neurosciences*, 30, 159–176.
- Hutto, D. D., & Myin, E. (2012). *Radicalizing enactivism: Basic minds without content*. Cambridge MA: MIT Press.
- Kallmeyer, L., & Osswald, R. (2013). Syntax-driven semantic frame composition in lexicalized tree adjoining grammars. *Journal of Language Modelling*, 1, 267–330.
- Kallmeyer, L., Osswald, R., & Pogodalla, S. (2017). Quantification in frame semantics with binders and nominals of hybrid logic. *Journal of Language Modelling*, 5, 357–383.
- Kant, I. (1781). *Kritik der reinen Vernunft*. Riga: Johann Friedrich Hartknoch.
- Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and Philosophy*, 4(1), 1–45.

- Krifka, M. (1990). Four thousand ships passed through the lock: Object-induced measure functions on events. *Linguistics and Philosophy*, 13, 487–520.
- Löbner, S. (2014). Evidence for frames from human language. In T. Gamerschlag, D. Gerland, W. Petersen, & R. Osswald (Eds.), *Frames and concept types (Studies in Linguistics and Philosophy 94)* (pp. 23–68). Heidelberg, New York: Springer.
- Löbner, S. (2015). Functional concepts and frames. In T. Gamerschlag, D. Gerland, R. Osswald, & W. Petersen (Eds.), *Meaning, frames, and conceptual representation* (pp. 15–42). Düsseldorf: Düsseldorf University Press.
- Löbner, S. (2020). The Partee paradox: Rising temperatures and numbers.' In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, & T. E. Zimmermann (Eds.), *The Wiley companion to semantics* (3 Vols.). Hoboken NJ: Wiley. <https://doi.org/10.1002/9781118788516.sem077>.
- Mackintosh, N. (1965). Selective attention in animal discrimination learning. *Psychological Bulletin*, 64, 124.
- Martin-Löf, P. (1984). *Intuitionistic type theory*. Napoli: Bibliopolis.
- Montague, R. (1970). Universal grammar. *Theoria*, 36, 373–398.
- Montague, R. (1973). The proper treatment of quantification in ordinary language. In J. Hintikka, J. Moravcsik, & P. Suppes (Eds.), *Approaches to natural language* (pp. 221–242). Dordrecht: Reidel.
- Montague, R. (1974). *Formal philosophy: Selected papers of Richard Montague*. New Haven CT: Yale University Press.
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge MA: The MIT Press.
- Partee, B. H. (1979). Semantics—mathematics or psychology? In R. Bäuerle, U. Egli, & A. von Stechow (Eds.), *Semantics from different points of view* (pp. 1–14). Berlin: Springer.
- Peacocke, C. (1992). *A study of concepts*. Cambridge MA: MIT Press.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6, 576–582.
- Pustejovsky, J. (1995). *The generative lexicon*. Cambridge MA: The MIT Press.
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale NJ: Lawrence Erlbaum Associates.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573–605.
- Smith, E. E., Osherson, D., Rips, L. J., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science*, 12, 485–527.
- Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning*. New York: Academic Press.
- Talmy, L. (2000). *Toward a cognitive semantics*. Vol. I: *Concept structuring systems*. Cambridge MA: MIT Press.
- Tarski, A. (1936). The concept of truth in formalised languages. Reprinted in Tarski (1956) *Logic, semantics, metamathematics. Papers from 1923–1938* (pp. 152–279). Oxford: Clarendon Press.
- Thomason, R. M. (Ed., 1974). *Formal philosophy. Selected papers of Richard Montague*. Edited and with an introduction by Richmond H. Thomason, New Haven and London: Yale University Press.
- Van Valin, R. D., Jr. & LaPolla, R. J. (1997). *Syntax*. Cambridge: Cambridge University Press.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# **Pushing the Boundaries of Formal Semantics**

# A Compositional Pluralist Semantics for Extensional and Attitude Verbs



Kristina Liefke

**Abstract** We propose a new account of linguistic content that reconciles content-pluralism with compositionality. This is achieved by integrating truth-conditional content and attitude report content into a single notion of content. A parametrized version of this notion (with parameters for agents, times, and information states) serves as input to the compositional semantic machinery. By supplying different parameter-values to the parametrized contents of their complements, different verbs select for different components of the complement's integrated content. The resulting account explains the different substitution properties of extensional and attitude constructions and captures the role of agents' epistemic perspective in the determination of attitude content. The account improves upon other accounts of truth-conditional and attitude content (esp. two-dimensional semantics) by interpreting different occurrences of an expression—in extensional and in attitude embeddings—as objects of *the same* semantic type, and by explaining the substitution-resistance of attitudinal embeddings of extensional constructions.

**Keywords** Pluralism about linguistic content · Compositional interpretation · Intensional verbs · Attitude reports · Epistemic perspective · Two-dimensional semantics

## 1 Introduction

The notion of linguistic content lies at the core of research in semantics and the philosophy of language. This notion describes the context-dependent meaning of (utterances of) linguistic expressions that is used to capture the truth-conditional contribution of these expressions and to predict the entailment relations between these expressions (see Lewis 1970; Montague 1970). Many semantic theories today adopt some form of pluralism about linguistic content (see, e.g., Zimmermann 2012;

---

K. Liefke (✉)

Institute for Linguistics, Goethe University Frankfurt, Norbert-Wollheim-Platz 1, 60323

Frankfurt am Main, Germany

e-mail: [Liefke@lingua.uni-frankfurt.de](mailto:Liefke@lingua.uni-frankfurt.de)

Ciardelli and Roelofsen 2018; Potts 2005). These theories assume different kinds, or types, of linguistic content that serve as the contents of expressions in different contexts and that, hence, play different explanatory roles.

Among the different kinds of linguistic content are typically *truth-conditional content* and *attitude (report)*<sup>1</sup> content. Truth-conditional content is sometimes alternatively called *denotational content*, *intensions*, or *objective meaning*. Attitude content is sometimes called *epistemic content*, *information content*, or *subjective meaning*. Respectively, these two kinds of content capture agent-independent criteria for assigning truth-values to utterances (i.e. truth-conditional content) and agents' particular ways of grasping the truth-conditional content of these utterances (i.e. attitude content).

The distinction between truth-conditional and attitude content is often motivated by the observation that certain linguistic constructions resist the truth-preserving substitution of truth-conditionally equivalent expressions in their complements. Such constructions include *de dicto*-readings of clausal embeddings under attitude verbs like *believe* or *hope*. Constructions that exhibit this substitution-resistance are called (*hyper-)*intensional constructions and can be described as *cognitively opaque*.<sup>2</sup> They differ from extensional<sup>3</sup> constructions (e.g. embeddings under the verb *indicate*) that allow for such substitutions and are, hence, *cognitively transparent*.

The difference between extensional and attitude constructions is reflected in the possibility, or impossibility, of substituting DPs like **sodium** by their co-referential DPs (here: **natrium**) and, hence, of substituting (1a) by the truth-conditionally equivalent (1b): while this substitution is typically allowed in the complement of *indicate* (s.t. one can infer (2b) from (2a)), it is often disallowed in the complement of *believe* (s.t. one cannot generally infer (3b) from (3a)). The latter inference is blocked if the attitude complements have a different cognitive significance for the attitude subject (in (3): for Len).

- (1) a. Sodium is a metal.
- b. Natrium is a metal.
- (2) a. The reaction indicates [<sub>CP</sub>that sodium is a metal]. (T)
- ⇒ b. The reaction indicates [<sub>CP</sub>that natrium is a metal]. (T)

<sup>1</sup>Because of our focus on *linguistic content*, we hereafter take attitude content to refer to the content of attitude *reports*, rather than to the content of the mental attitudes underlying these reports (see Hintikka 1969).

<sup>2</sup>Our notion of *cognitive opacity* differs from the familiar notion of (referential) opacity (see Quine 1953), which captures the sensitivity for truth-conditional, rather than for attitude content. Our notion of *cognitive transparency* differs from referential transparency, which captures the sensitivity for reference/extension. The difference between these notions is exemplified by the verb *indicate*, which creates a referentially opaque, but cognitively transparent context.

<sup>3</sup>We will hereafter use *extensional verb* (or *construction*) as a cover term for verbs (or constructions) that take *extensional* and for verbs (or constructions) that take *intensional* complements. Our use of this term is motivated by the common description of objectual attitude verbs as *intensional transitive verbs*.

- (3) a. Len believes [<sub>CP</sub>that sodium is a metal]. (T)  
**⇒ b.** Len believes [<sub>CP</sub>that sodium is a metal]. (F)

To explain the difference in substitutivity between (2) and (3), most pluralist theories about linguistic content (e.g. Chalmers 2006; Zimmermann 2012; Lappin 2015) interpret extensional verbs as expressions that select for the *truth-conditional* content of their complement and interpret attitude verbs as expressions that select for the *combined* (truth-conditional and *attitude*) content of their complement. However, these theories often yield a disunified semantics that interprets different occurrences of a complement—in extensional and in attitude embeddings—as objects of different types. As a result, these theories often resist an easy compositional formulation. However, given their intended role as an account of natural language content, this is highly problematic.

This paper outlines a new account of truth-conditional and attitude content, called *Integrated Semantics*, that solves the above problem by integrating truth-conditional and attitude content into a single notion of linguistic content. The account enables a uniform compositional treatment of extensional and attitude constructions that correctly predicts the substitution behavior of these constructions.

The paper is organized as follows: To show the need for an integrated account of truth-conditional and attitude content, we first describe the relation between truth-conditional and attitude content, review the most popular account of these two kinds of content (i.e. two-dimensional semantics), and identify some shortcomings of this account (in Sect. 2). The rest of the paper will be concerned with an incremental presentation of our alternative account of truth-conditional and attitude content, i.e. *Integrated Semantics*, and with a demonstration of the ability of this account to avoid the above shortcomings. To this aim, we first give an informal presentation of Integrated Semantics (in Sect. 3), which we subsequently turn into a compositional semantics for a small fragment of English containing extensional and attitude verbs (in Sect. 4, 5). The paper closes with a summary of our results and with pointers to future work.

## 2 Accounts of Truth-Conditional and Attitude Content

The distinction between truth-conditional and attitude content is anticipated by the different roles of Frege’s notion of *sense* [German *Sinn*]. In (Frege 1892), the sense of an expression serves both to determine the denotation [*Bedeutung*] of this expression and to provide the linguistic content of this expression in indirect (e.g. attitude) contexts. The latter role is enabled by the fact that the sense of an expression contains the denotation’s mode of presentation [*Art des Gegebenseins*; MoP] to the cognitive agent. Newer work in semantics captures the difference between the above roles by distinguishing, e.g., between truth-conditional content/reference and *guises* of this

content (see Heim 1998), between truth-conditional content and epistemic *roles* (see Perry 2012), between intensions and *intentions* (see Thomason 1980), and between objects and cognitive *concepts* (see Barsalou 1992).

The distinction between truth-conditional and attitude content is sometimes also captured by separating *hyperintensions* from Carnapian *intensions*: intensions of linguistic expressions are functions from indices (i.e. worlds, or world-time pairs) to the expressions' denotations at these indices (see Carnap 1988; Montague 1970). Intensions thus encode the expressions' truth-conditional content. Hyperintensions are objects with stricter identity-conditions than intensions that serve as the complements of attitude verbs, i.e. they play the role of attitude content. Hyperintensions typically take the form of structured contents (see Lewis 1970; Cresswell 1985), of sets of (im-)possible worlds/situations (see Muskens 1995; Zalta 1997), of unanalyzable primitives (see Thomason 1980; Pollard 2015), or of computational operations (see Moschovakis 2006; Lappin 2015).

## 2.1 *The Relation Between Truth-Conditional and Attitude Content*

Most theories of linguistic content assume some relation between truth-conditional and attitude content. This relation is suggested by Frege's assumption that the sense of an expression (*qua* MoP) determines the expression's denotation. The possibility of obtaining truth-conditional content from attitude content enables a compositional semantics for extensional and attitude verbs. However, this possibility is compromised by the fact that speakers' actual MoPs often underdetermine or misdetermine the expression's denotation. In particular, Kripke (1980) has observed that speakers often lack uniquely identifying information about the expression's denotation (s.t. their MoPs identify *other objects in addition to* the expression's denotation) or have false information about this denotation (s.t. their MoPs identify *a different object than* the expression's denotation).

To avoid the challenge from under- or misdetermination, many contemporary theories treat truth-conditional content as the 'default' kind of content and only introduce attitude content in response to special contextual triggers (e.g. occurrence in the complement of an attitude verb). However, this strategy causes a serious problem for the compositional interpretation of natural language: to enable the compositional interpretation of attitude reports, the linguistic content of the attitude complement (i.e. an *attitude* content) must, in some way, be obtainable from the kind of content that serves as input to the compositional machinery (here: a *truth-conditional* content). However, since attitude content is often richer than truth-conditional content, this is not generally possible.

## 2.2 Attempts at (Re-)Connecting Truth-Conditional and Attitude Content

In semantics and the philosophy of language, there have been some recent efforts towards a theory of truth-conditional and attitude content that avoids the dilemma between under- or misdetermination and non-compositionality. These efforts include two-dimensional semantics (see Kaplan 1989; Haas-Spohn 1995; Chalmers 2006; Zimmermann 2012), which interprets linguistic expressions as functions (Kaplanian *characters*) from contexts to intensions, i.e. as functions from contexts to *contents*. Contexts  $c$  are tuples containing the world  $w_c$ , time  $t_c$ , location  $l_c$ , and agent/speaker  $a_c$  of the context. Intensions are functions from indices to extensions. The intension of a character  $\chi$  at a context  $c$ ,  $\lambda w \lambda t. \chi(c)(w, t)$ , serves the role of truth-conditional content. The diagonal of a character  $\chi$ , i.e. a function,  $\lambda c. \chi(c)(w_c, t_c)$ , from contexts to the character's extension at the context and the context's index,  $\langle w_c, t_c \rangle$ , serves the role of attitude content.

Two-dimensional semantics has been a remarkable success story. However, this semantics faces several problems regarding the compositional interpretation of attitude reports. These problems are identified below. We will see that each of these problems motivates a desideratum for an alternative, compositional theory of *integrated* (truth-conditional and attitude) content.

### 2.2.1 Problem 1: Empirical Adequacy

To explain the substitution behavior of attitude reports (see (3)), most theories of two-dimensional semantics (e.g. Lerner and Zimmermann 1991; Haas-Spohn 1995; Schlenker 2003) treat proper names and kind terms as indexical expressions whose truth-conditional content is determined by the utterance context. In virtue of this treatment, co-referential names/kind terms are assigned different characters. The interpretation of attitude verbs as relations to *characters* (or to *diagonals* of characters) and the identification of compositionality with compositionality of character<sup>4</sup> then explain the substitution failure in (3). However, without further—still underexplored—restrictions on the notion of character, the resulting semantics gives *trivial, inadequate truth-conditions* for attitude reports (see von Stechow and Zimmermann 2004).

---

<sup>4</sup> According to this principle, the character of a complex expression is a function of the characters of the expression's syntactic constituents and their mode of composition (see Westerståhl 2012). The adoption of this principle predicts the preservation of an expression's character under the substitution of same-character constituents.

### 2.2.2 Problem 2: Semantic Uniformity

To capture the different substitution properties of extensional and attitude constructions (e.g. (2) vs. (3)), some two-dimensional theories (esp. Chalmers 2006; see Lerner and Zimmermann 1991) vary the interpretation of expressions with the expressions' linguistic context: when an expression occurs in the complement of an attitude verb, it is interpreted as the *diagonal* of its character; otherwise, it is interpreted as its intension. However, this variation challenges the uniform interpretation of extensional verbs: since constructions like (2a) often lose their cognitive transparency in attitude embeddings (note the cognitive difference-for-Len between (1a) and (1b), and the resulting non-substitutivity of (2a) by (2b) in (4a)), extensional verbs require—next to their ‘extensional’ interpretation (on which they take intension-type complements)—a hyperintensional interpretation (on which they take diagonal-type complements). But this doubling seriously complicates their compositional interpretation (cf. Theiler et al. 2018; Liefke and Werning 2018).

- (4) a. Len believes [that the reaction indicates [that sodium is a metal]]. (T)  
 $\Rightarrow$  b. Len believes [that the reaction indicates [that sodium is a metal]]. (F)

### 2.2.3 Problem 3: Perspective-Dependence

The treatment of attitude reports in two-dimensional semantics is further challenged by the inability of this semantics to explain agent- and time-specific differences in the substitutivity of truth-conditionally equivalent complements (compare (3) and (5)). To account for these differences, some two-dimensional theories (e.g. Haas-Spohn 1995) relativize the diagonal of an attitude complement to the attitude subject (i.e. to the object at the origin of the causal chain of uses of the complement's name-constituent in the subject's language). However, apart from the need for further relativization (e.g. to the *time* of use; see the difference in substitutivity between (3)' and (6), which assumes the cognitive identity-for-Len of (1a) and (1b) at the later point in time  $t_{k+1}$ ), it is not clear how this relativization can be implemented in a *compositional* interpretation of attitude reports.

- (5) a. Eve believes [<sub>CP</sub>that sodium is a metal]. (T)  
 $\Rightarrow$  b. Eve believes [<sub>CP</sub>that sodium is a metal]. (T)!  
(3)' a. Len believes (at  $t_k$ ) [<sub>CP</sub>that sodium is a metal]. (T)  
 $\Rightarrow$  b. Len believes (at  $t_k$ ) [<sub>CP</sub>that sodium is a metal]. (F)  
(6) a. Len believes (at  $t_{k+1}$ ) [<sub>CP</sub>that sodium is a metal]. (T)  
 $\Rightarrow$  b. Len believes (at  $t_{k+1}$ ) [<sub>CP</sub>that sodium is a metal]. (T)!

### 2.3 Desiderata for an Account of Truth-Conditional and Attitude Content

The above problems suggest an alternative theory of truth-conditional and attitude content that has the following properties:

- (P.1) The theory is *compositional* [cf. the problem of empirical adequacy]
- (P.2) The theory gives *adequate truth- and entailment-conditions* for extensional and attitude constructions [cf. the problem of empirical adequacy]
- (P.3) The theory enables a *uniform interpretation* of extensional and attitude constructions [cf. the problem of semantic uniformity]
- (P.4) The theory *accommodates agents' epistemic perspective* on the entities in the domain of discourse [cf. the problem of perspective-dependence]

At present, there does not exist a theory of linguistic content that satisfies all of (P.1) to (P.4). However, such a theory is essential for the adequate compositional interpretation of natural language.

## 3 Integrated Semantics

*Integrated Semantics* [hereafter, IS] is a novel account of linguistic content that satisfies properties (P.1) to (P.4). This account is a version of two-dimensional semantics that obtains linguistic contents by applying meanings to contexts (here: to centered informational situations). In contrast to *contents* in two-dimensional semantics, contents in Integrated Semantics contain attitude content next to their familiar truth-conditional content. We call the relevant notion of content *integrated content*, abbreviated 'IC'. A parametrized version of this notion (with a parameter for centered informational situations; dubbed 'parametrized IC', or 'PIC') serves as input to the compositional semantic machinery. By supplying different centered situations to the PICs of their complements, different verbs select for different (truth-conditional, or integrated) components of their complement's IC. This selection explains the distinct substitution behaviour of the verbs' complements.

Below, we first introduce centered (informational) situations (in Sect. 3.1). We then give an initial presentation of IS. This presentation proceeds by describing the IC of sentences and proper names at a centered situation (in Sects. 3.2, 3.3).

### 3.1 Centered Informational Situations

Centered informational situations (or simply, *centered situations*) are ordered triples  $\sigma^* := \langle \sigma, t_\sigma, a_\sigma \rangle$  consisting of an informational situation  $\sigma$ , a point in time  $t_\sigma$ , and

a cognitive agent  $a_\sigma$ .<sup>5</sup> Such triples represent the informational situation of  $a_\sigma$  at  $t_\sigma$ . Because of our particular use of such situations, we do not require that  $\sigma$  contains information about  $a_\sigma$  him-/herself.

Informational situations  $\sigma$  are world-level<sup>6</sup> correlates of information states. Such states are typically represented by sets of worlds (i.e. sets of those worlds that are compatible with the available information in this state). In virtue of the correspondence between situations and information states, every sentence that is true (or false) at all worlds in an information state is true (resp. false) in the corresponding situation. This is made possible by the partiality of situations: a sentence may be neither true nor false in a situation. The partiality of situations captures the informational imperfection of cognitive agents. To allow for the possibility of false information, we also consider impossible situations (see Zalta 1997).

The partial nature of informational situations induces a partial ordering on the set of situations. In particular, a situation  $\sigma_2$  *includes* a situation  $\sigma_1$  if  $\sigma_2$  contains all information that is contained in  $\sigma_1$ . We call any situation that includes a situation an *extension* of that situation and identify the *maximal (consistent) extension* of a situation with a (possible) world extending this situation. We assume that every ordering of situations has a bottom element (called the ‘empty’ situation; denoted ‘†’) and a top element (some world  $w$ ). We assume a single empty situation.

As a consequence of the correspondence between informational situations and sets of worlds, situations have fairly coarse-grained identity conditions. For example, sentences that contain different co-referential or truth-conditionally equivalent expressions (e.g. (1a), (1b)) are true (or false) in the same situations. The ‘enrichment’ of informational situations by cognitive agents and points in time compensates for this shortcoming, as we will see below.

### 3.2 The Integrated Content of Sentences

We have mentioned above that a sentence’s integrated content at a centered situation contains both truth-conditional and attitude content. To combine these two kinds of content into a single notion of ‘integrated’ content, Integrated Semantics identifies the integrated content of a sentence with the result of restricting the sentence’s classical truth-conditional content at a centered situation (i.e. the set of worlds or situations in which the sentence is true) to smaller sets of situations that also encode the interpreter’s salient description, guise, or MoP of the sentence’s constituents at the time of interpretation. For (1a) and the centered situation  $\sigma_0^* := \langle \sigma_0, t, a \rangle$  (where  $a$  is the sentence’s interpreter), such a set is given in (7).

In what follows, we will use denotation brackets,  $\llbracket \cdot \rrbracket$ , as a notational device for the IS-interpretation of linguistic expressions. The PIC of the sentence **Sodium** is

<sup>5</sup>Centered informational situations are, thus, a variant of centered situations (see Stephenson 2010), which are ordered pairs of an agent and a world-part.

<sup>6</sup>Situations are thus objects of type  $s$ , not of the type of information states,  $\langle s, t \rangle$ .

a metal (i.e. (1a)) is then denoted by ‘[Sodium is a metal]’. The IC of this sentence at the centered situation  $\sigma_0^*$  is denoted by ‘[Sodium is a metal] $(\sigma_0^*)$ ’ (see (7)). In (7),  $\llbracket \text{sodium} \rrbracket(\sigma_0^*)$  is the set of properties that captures  $a$ 's MoP of sodium in  $\sigma_0$  at  $t$ . This set is obtained from the IS-interpretation of the name sodium at  $\sigma_0^*$  (see Sect. 3.3) and enters the IC of (1a) through the sentence's compositional interpretation at  $\sigma_0^*$  (see Sect. 4). Below, we use  $\sigma$  as a variable over situations.

$$\begin{aligned} & \llbracket \text{Sodium is a metal} \rrbracket(\sigma_0^*) \\ = & \{\sigma \mid \underbrace{\text{sodium is a metal in } \sigma}_{\text{truth-cond'l content}} \& \underbrace{\text{sodium has all properties from } \llbracket \text{sodium} \rrbracket(\sigma_0^*) \text{ in } \sigma}_{\text{attitude content (at } \sigma_0^*)}\} \end{aligned} \quad (7)$$

As a result of the coarse grain of situations (in particular, by the identification of sodium- and sodium- (i.e. Na-)containing situations), (7) is equivalent to (8):

$$\{\sigma \mid \text{Na is a metal in } \sigma \& \text{Na has all properties from } \llbracket \text{sodium} \rrbracket(\sigma_0^*) \text{ in } \sigma\} \quad (8)$$

The first restriction on the set from (7) (see the grey underbrace) identifies the *truth-conditional content* of (1a). The second restriction (see the black underbrace) identifies the *attitude content* of (1a) at  $a$ 's information state  $\sigma_0$  at time  $t$ . Since truth-conditional and attitude content perform different restrictions on the same set of situations, a sentence's IC is an object of the same type (i.e. a set of situations, type  $\langle s, t \rangle$ ) as the truth-conditional and the attitude component of this IC. This enables the same-type interpretation of the occurrences of the verb indicate in (2a) and (4a). Integrated Semantics thus meets Desideratum (P.3).

Notably, by integrating an expression's (*agent-independent*) truth-conditional content with its (*agent-dependent*) attitude content, we do not suggest that linguistic agents know the expression's truth-conditional content: the agentive center of the situation  $\sigma_0^*$  may possess the information contained in (1a)'s attitude content at  $\sigma_0^*$  *without* thereby also possessing the information contained in (1a)'s truth-conditional content. For example,  $a$  may be unaware of the referential relation between the name sodium and the chemical element Na. In (7), the element Na only provides an ‘external anchor’ for the properties in the set  $\llbracket \text{sodium} \rrbracket(\sigma_0^*)$ . While this anchor simplifies the representation of integrated content, nothing depends on it.

### 3.3 The Interpretation of Proper Names

In Integrated Semantics, proper names (e.g. sodium) are interpreted as intensional generalized quantifiers [IQs], i.e. as functions from centered situations to partial sets of properties of individuals. This interpretation is justified by the existence of a non-injective function,  $\circ$ , from IQs to individuals, s.t. we can obtain the referent of a name from the name's PIC. The non-injective nature of this function captures the intuitive semantic distinctness of co-referential names.

We illustrate the IS-interpretation of names through an example: assume that, in  $\sigma_1$  at  $t_7$ , Len thinks of sodium as the reactive substance and of sodium as the silvery-white substance and that, in  $\sigma_4$  at  $t_7$ , Eve thinks of both sodium and sodium as the silvery-white reactive metal. The IQs,  $\llbracket \text{sodium} \rrbracket$  and  $\llbracket \text{natruium} \rrbracket$ , that serve as the PICs of the names **sodium** and **natrium**, then have the following values at  $\sigma_{len}^* := \langle \sigma_1, t_7, len \rangle$  and  $\sigma_{eve}^* := \langle \sigma_4, t_7, eve \rangle$ :

$$\llbracket \text{sodium} \rrbracket(\sigma_{len}^*) = \{\text{is reactive}\} \quad (9a)$$

$$\llbracket \text{natruium} \rrbracket(\sigma_{len}^*) = \{\text{is silvery-white}\} \quad (9b)$$

$$\llbracket \text{sodium} \rrbracket(\sigma_{eve}^*) = \{\text{is reactive, is silvery-white, is a metal}\} \quad (9c)$$

$$= \llbracket \text{natruium} \rrbracket(\sigma_{eve}^*)$$

On the basis of the above, (1a) and (1b) are interpreted as (10) and (11) by Len, and as (12) by Eve:

$$\begin{aligned} & \llbracket \text{Sodium is a metal} \rrbracket(\sigma_{len}^*) && (10) \\ = & \{\sigma \mid \text{Na is a metal in } \sigma \& \text{Na has all properties from } \llbracket \text{sodium} \rrbracket(\sigma_{len}^*) \text{ in } \sigma\} \\ = & \{\sigma \mid \text{Na is a metal in } \sigma \& \text{Na is reactive in } \sigma\} \\ \neq & \{\sigma \mid \text{Na is a metal in } \sigma \& \text{Na is silvery-white in } \sigma\} \\ = & \{\sigma \mid \text{Na is a metal in } \sigma \& \text{Na has all properties of } \llbracket \text{natruium} \rrbracket(\sigma_{len}^*) \text{ in } \sigma\} \\ = & \llbracket \text{Natrium is a metal} \rrbracket(\sigma_{len}^*) && (11) \end{aligned}$$

$$\begin{aligned} & \llbracket \text{Sodium is a metal} \rrbracket(\sigma_{eve}^*) && (12) \\ = & \{\sigma \mid \text{Na is a metal in } \sigma \& \text{Na has all properties of } \llbracket \text{sodium} \rrbracket(\sigma_{eve}^*) \text{ in } \sigma\} \\ = & \{\sigma \mid \text{Na is a metal in } \sigma \& \text{Na is silvery-white and reactive in } \sigma\} \\ = & \{\sigma \mid \text{Na is a metal in } \sigma \& \text{Na has all properties of } \llbracket \text{natruium} \rrbracket(\sigma_{eve}^*) \text{ in } \sigma\} \\ = & \llbracket \text{Natrium is a metal} \rrbracket(\sigma_{eve}^*) \end{aligned}$$

The *difference* between the ICs of (1a) and (1b) at  $\sigma_{len}^*$  – and their *identity* at  $\sigma_{eve}^*$  – captures Len’s and Eve’s different epistemic perspectives on the referents of **sodium** and **natrium**, and explains the difference in substitutivity between (3) and (5). As a result, Integrated Semantics also meets Desiderata (P.2) and (P.4).

## 4 The Compositional Interpretation of VPs

We have suggested above that the attitude content of (1a) at  $\sigma_0^*$  is obtained from the value-at- $\sigma_0^*$  of the IS-interpretation of the name **sodium**. The present section specifies the interpretation of the VP **is a metal**, which obtains the IC of (1a) from this value. To keep this specification as simple as possible—and to make the interpretation

of linguistic expressions reminiscent to the description of sentence-interpretations from the previous section –, we combine set-theoretic with lambda notation.<sup>7</sup> In the resulting ‘mixed’ notation, the PIC of (1a) is described as follows (cf. (8)):

$$\begin{aligned} & \llbracket \text{Sodium is a metal} \rrbracket \\ = & \lambda \sigma^*. \{ \sigma \mid \text{Na is a metal in } \sigma \text{ \& Na has all properties from } \llbracket \text{sodium} \rrbracket(\sigma^*) \text{ in } \sigma \} \end{aligned} \quad (13)$$

We have mentioned in the previous section that the PICs of names are related to the names’ individual referents through the non-injective function  $\circ$ . This function allows us to render (13) as (14), where  $\circ$  is written in postfix notation (s.t. ‘ $x^\circ$ ’ denotes  $\circ(x)$ ):

$$\begin{aligned} & \lambda \sigma^*. \{ \sigma \mid \llbracket \text{sodium} \rrbracket^\circ \text{ is a metal in } \sigma \text{ \&} \\ & \llbracket \text{sodium} \rrbracket^\circ \text{ has all properties from } \llbracket \text{sodium} \rrbracket(\sigma^*) \text{ in } \sigma \} \end{aligned} \quad (14)$$

Axiom **Ax1** ensures the non-injectivity of  $\circ$ . Below, we let  $x$  and  $y$  be variables over IQs.

$$\exists x \exists y [x^\circ = y^\circ \wedge x \neq y] \quad (\mathbf{Ax1})$$

**Ax1** is instantiated by the relation between the PICs of **sodium** and **natrium** (in a standard model, given a standard interpretation function):

$$\text{Na} = \llbracket \text{sodium} \rrbracket^\circ = \llbracket \text{natrium} \rrbracket^\circ \wedge \llbracket \text{sodium} \rrbracket \neq \llbracket \text{natrium} \rrbracket \quad (15)$$

The PICs of the name **sodium** and of sentence (1a) (cf. (14)) then suggest the following interpretation of the VP **be a metal** (in (16)): (For simplicity, we treat this VP as a single lexical unit.)

$$\begin{aligned} & \llbracket \text{be a metal} \rrbracket \\ = & \lambda x \lambda \sigma^*. \{ \sigma \mid x^\circ \text{ is a metal in } \sigma \text{ \& } x^\circ \text{ has all properties from } x(\sigma^*) \text{ in } \sigma \} \end{aligned} \quad (16)$$

The above enables the compositional interpretation of (1a) at  $\sigma_0^*$  as follows:

$$\begin{aligned} & \llbracket [\text{DP Sodium}] [\text{VP is a metal}] \rrbracket(\sigma_0^*) \\ = & \lambda x \lambda \sigma^*. \{ \sigma \mid x^\circ \text{ is a metal in } \sigma \text{ \&} \\ & x^\circ \text{ has all properties from } x(\sigma^*) \text{ in } \sigma \} \{ \llbracket \text{sodium} \rrbracket \}(\sigma_0^*) \\ \equiv & \lambda \sigma^*. \{ \sigma \mid \llbracket \text{sodium} \rrbracket^\circ \text{ is a metal in } \sigma \text{ \&} \\ & \llbracket \text{sodium} \rrbracket^\circ \text{ has all properties from } \llbracket \text{sodium} \rrbracket(\sigma^*) \text{ in } \sigma \}(\sigma_0^*) \\ \equiv & \lambda \sigma^*. \{ \sigma \mid \text{Na is a metal in } \sigma \text{ \& Na has all properties from } \llbracket \text{sodium} \rrbracket(\sigma^*) \text{ in } \sigma \}(\sigma_0^*) \\ = & \{ \sigma \mid \text{Na is a metal in } \sigma \text{ \& Na has all properties from } \llbracket \text{sodium} \rrbracket(\sigma_0^*) \text{ in } \sigma \} \end{aligned} \quad (17)$$

---

<sup>7</sup>The resulting ‘mixed’ notation is adopted, e.g., in (Ciardelli et al. 2017).

With the interpretation of names and VPs in place, we next turn to the interpretation of clausally complemented verbs in Integrated Semantics.

## 5 Extensional and Attitude Verbs in IS

We have seen in Sect. 1 that different clausally complemented verbs impose differently strong restrictions on the substitutivity of their complements. Integrated Semantics captures this difference by assuming that different verbs supply different centered situations to the PICs of their complements.<sup>8</sup> In particular, while extensional verbs like **indicate** typically<sup>9</sup> supply a designated centered situation (hereafter called *the ‘empty’ centered situation*, denoted by ‘†\*’) that contains the empty situation †, attitude verbs like **believe** supply a contextually chosen centered situation that depends on the particular state or event described by the verb. Below, we first describe the interpretation of extensional verbs in IS (in Sect. 5.1). We then turn to the interpretation of attitude verbs (in Sect. 5.2) and of attitudinal embeddings of extensional verbs (in Sect. 5.3).

### 5.1 The Interpretation of Extensional Verbs

In Sect. 3.1, we have identified the ‘empty’ situation † as the bottom element in the partial ordering on situations, at which no sentence is true or false. As a result of this characterization, the set of properties that is associated with the name **sodium** at the centered situation †\* will be empty. This is captured in **Ax2**. Below,  $x$  and  $P$  are variables over IQs and properties, respectively.

$$\forall x [x(\dagger^*) = (\lambda P. \perp)] \quad (\mathbf{Ax2})$$

The interpretation of the verb **indicate** is given below, where  $p$  is a variable over PICs<sup>10</sup>:

$$\llbracket \text{indicate} \rrbracket = \lambda p \lambda x \lambda \sigma^*. \{ \sigma \mid x^\circ \text{ indicates } p(\dagger^*) \text{ in } \sigma \} \quad (18)$$

The above interpretation enables the compositional interpretation of (2a) at  $\sigma_0^*$  as follows:

---

<sup>8</sup>Since their interpretation thus influences the content of their complement, such verbs are Kaplanian *monsters* (see Kaplan 1989, Sect. VIII). The ‘monstrous’ interpretation of attitude verbs follows (Israel and Perry 1996) and (Schlenker 2003).

<sup>9</sup>This is not the case in attitudinal embeddings of such verbs, as we show in Sect. 5.3.

<sup>10</sup>In order to allow its application to the entire sentence, this interpretation stipulates a simplistic semantics for the DP **the reaction** (see Sect. 6).

$$\begin{aligned}
 & [[\text{the reaction} [\text{indicates} [\text{that sodium is a metal}]]]](\sigma_0^*) & (19) \\
 = & \lambda p \lambda x \lambda \sigma^*. \{\sigma \mid x^\circ \text{ indicates } p(\dagger^*) \text{ in } \sigma\} (\lambda \sigma^*. \{\sigma' \mid \text{Na is a metal in } \sigma' \& \\
 & \text{Na has all p'ties from } [[\text{sodium}](\sigma_0^*) \text{ in } \sigma'])([[\text{the reaction}]])(\sigma_0^*) \\
 = & \lambda \sigma^*. \{\sigma \mid \text{the reaction indicates } \{\sigma' \mid \text{Na is a metal in } \sigma' \& \\
 & \text{Na has all properties from } [[\text{sodium}](\dagger^*) \text{ in } \sigma'] \text{ in } \sigma\}(\sigma_0^*) \\
 = & \{\sigma \mid \text{the reaction indicates } \{\sigma' \mid \underbrace{\text{Na is a metal in } \sigma'}_{\text{truth-cond'l content}} \& \\
 & \underbrace{\text{Na has all properties from } [[\text{sodium}](\dagger^*) \text{ in } \sigma']}_{\text{attitude content (at } \dagger^*)} \text{ in } \sigma\} \\
 = & \{\sigma \mid \text{the reaction indicates } \{\sigma' \mid \underbrace{\text{Na is a metal in } \sigma'}_{\text{truth-cond'l content}} \text{ in } \sigma\}
 \end{aligned}$$

The above shows that the application of the IS-interpretation of the complement of **indicate** to the empty centered situation effectively *deletes* the attitude content of the complement. This reflects the fact that extensional verbs only select for the *truth-conditional* component of their complement. As a result of this selection, (2b) has the same PIC (and, hence, the same IC-at- $\sigma_0^*$ ) as (2a) (see (20)), such that the former can be substituted *salva veritate* for the latter.

$$\begin{aligned}
 & [[\text{the reaction} [\text{indicates} [\text{that sodium is a metal}]]]](\sigma_0^*) & (20) \\
 = & \{\sigma \mid \text{the reaction indicates } \{\sigma' \mid \text{Na is a metal in } \sigma' \& \text{Na has all} \\
 & \text{properties from } [[\text{natrium}](\dagger^*) \text{ in } \sigma'] \text{ in } \sigma\} \\
 = & \{\sigma \mid \text{the reaction indicates } \{\sigma' \mid \text{Na is a metal in } \sigma'\} \text{ in } \sigma\}
 \end{aligned}$$

## 5.2 The Interpretation of Attitude Verbs

In contrast to extensional verbs, attitude verbs obtain their complement's IC at a centered situation that is provided by a pragmatically given choice function (see von Heusinger 2013). This function selects, from the set of all centered situations,  $\Sigma^*$ , a centered situation whose situation-coordinate the ascriber of the attitude ascribes to the bearer of the attitude at the time of the ascription.

Since the attitude ascriber and the ascription-time are coordinates in the centered situation at which the attitude report is interpreted (hereafter, *the external (centered) situation*), the choice of the ascribed situation (i.e. of the *internal* (centered) situation) depends on the external situation. Since the standards of information vary with different attitudes (e.g. knowledge vs. belief), the choice of situation further depends on the particular state or event that is described by the attitude verb. Below, we represent these dependencies by superscripting the constant,  $f$ , for the choice function with the external situation, and by co-indexing this constant with the attitude verb. The resulting interpretation of the verb **believe** is given in (21).

$$\llbracket \text{believe} \rrbracket = \lambda p \lambda x \lambda \sigma^*. \{ \sigma \mid x^\circ \text{ believes}^i p(f_i^{\sigma^*}(\Sigma^*)) \text{ in } \sigma \} \quad (21)$$

The compositional interpretation of (3a) at  $\sigma_0^*$  is given below:

$$\begin{aligned} & \llbracket \text{Len [believes [that sodium is a metal]]} \rrbracket(\sigma_0^*) \\ &= \lambda p \lambda x \lambda \sigma^*. \{ \sigma \mid x^\circ \text{ believes}^i p(f_i^{\sigma^*}(\Sigma^*)) \text{ in } \sigma \} (\lambda \sigma^*. \{ \sigma' \mid \text{Na is a metal in } \sigma' \& \text{ has all properties from } \llbracket \text{sodium} \rrbracket(\sigma_0^*) \text{ in } \sigma' \})(\llbracket \text{Len} \rrbracket)(\sigma_0^*) \\ &= \lambda \sigma^*. \{ \sigma \mid \text{Len believes}^i \{ \sigma' \mid \text{Na is a metal in } \sigma' \& \text{ Na has all properties from } \llbracket \text{sodium} \rrbracket(f_i^{\sigma^*}(\Sigma^*)) \text{ in } \sigma' \} \text{ in } \sigma \}(\sigma_0^*) \\ &= \{ \sigma \mid \text{Len believes}^i \{ \sigma' \mid \underbrace{\text{Na is a metal in } \sigma'}_{\text{truth-cond'l content}} \& \\ &\qquad \underbrace{\text{Na has all p'ties from } \llbracket \text{sodium} \rrbracket(f_i^{\sigma_0^*}(\Sigma^*)) \text{ in } \sigma'}_{\text{attitude content (at } f_i^{\sigma_0^*}(\Sigma^*))} \} \text{ in } \sigma \} \end{aligned} \quad (22)$$

Assume that  $\sigma_0^*$  has as its agentive center an *accurate* attitude ascriber, such that  $f_i^{\sigma_0^*}(\Sigma^*) = \sigma_{len}^*$  for  $i$  the index of **Len believes**, and  $f_j^{\sigma_0^*}(\Sigma^*) = \sigma_{eve}^*$  for  $j$  the index of **Eve believes** (see Sect. 3.3). Then, the pairs of sentences from (3) and (5) are interpreted as (23) and (24), and as (25), respectively:

$$\begin{aligned} & \llbracket \text{Len [believes [that sodium is a metal]]} \rrbracket(\sigma_0^*) \\ &= \{ \sigma \mid \text{Len believes}^i \{ \sigma' \mid \text{Na is a metal in } \sigma' \& \\ &\qquad \text{Na has all p'ties from } \llbracket \text{sodium} \rrbracket(f_i^{\sigma_0^*}(\Sigma^*)) \text{ in } \sigma' \} \text{ in } \sigma \} \\ &= \{ \sigma \mid \text{Len believes} \{ \sigma' \mid \text{Na is a metal in } \sigma' \& \\ &\qquad \text{Na has all properties from } \llbracket \text{sodium} \rrbracket(\sigma_{len}^*) \text{ in } \sigma' \} \text{ in } \sigma \} \\ &= \{ \sigma \mid \text{Len believes} \{ \sigma' \mid \text{Na is a metal in } \sigma' \& \text{Na is reactive in } \sigma' \} \text{ in } \sigma \} \\ &\neq \{ \sigma \mid \text{Len believes} \{ \sigma' \mid \text{Na is a metal in } \sigma' \& \text{Na is silvery-white in } \sigma' \} \text{ in } \sigma \} \\ &= \llbracket \text{Len [believes [that sodium is a metal]]} \rrbracket(\sigma_0^*) \end{aligned} \quad (23)$$

$$\begin{aligned} & \llbracket \text{Eve [believes [that sodium is a metal]]} \rrbracket(\sigma_0^*) \\ &= \{ \sigma \mid \text{Eve believes}^j \{ \sigma' \mid \text{Na is a metal in } \sigma' \& \\ &\qquad \text{Na has all p'ties from } \llbracket \text{sodium} \rrbracket(f_j^{\sigma_0^*}(\Sigma^*)) \text{ in } \sigma' \} \text{ in } \sigma \} \\ &= \{ \sigma \mid \text{Eve believes} \{ \sigma' \mid \text{Na is a metal in } \sigma' \& \\ &\qquad \text{Na has all properties from } \llbracket \text{sodium} \rrbracket(\sigma_{eve}^*) \text{ in } \sigma' \} \text{ in } \sigma \} \\ &= \{ \sigma \mid \text{Eve believes} \{ \sigma' \mid \text{Na is a metal, silvery-white, and reactive in } \sigma' \} \text{ in } \sigma \} \\ &= \llbracket \text{Eve [believes [that sodium is a metal]]} \rrbracket(\sigma_0^*) \end{aligned} \quad (24)$$

The above shows that—in contrast to the verb **indicate**—**believe** does not, in general, allow the truth-preserving substitution of truth-conditionally equivalent CPs in its complement. This is due to the fact that the internal situation at which the complement's IC is obtained preserves the attitude content of the complement of **believe** (see the black underbrace in (22)). As a result, the substitutivity of equivalent CPs only holds, in general, for CPs that have the same IC at all centered situations and, specifically, for CPs that also have the same attitude content at the particular centered situation at which the complement's IC is obtained. The latter case explains bearer-(and ascriber-)specific differences in the substitutivity of equivalent complements of attitude reports (see (P.4)).

### 5.3 Attitudinal Embeddings of Extensional Verbs

The interpretation of extensional and attitude verbs from the last two subsections enables the compositional interpretation of constructions containing these verbs (s.t. Integrated Semantics also meets Desideratum (P.1)). However, the interpretation of extensional complements at the situation  $\dagger^*$  (see Sect. 5.1) fails to capture the substitution-resistance of truth-conditionally equivalent complements of extensional verbs that occur in attitude embeddings (see (4)). To compensate for this shortcoming, we also interpret the complements of extensional verbs at a contextually given centered situation. The IS-interpretation of the verb **indicate** from (18) is then replaced by the interpretation below:

$$\llbracket \text{indicate} \rrbracket = \lambda p \lambda x \lambda \sigma^*. \{ \sigma \mid x^\circ \text{ indicates}^i p (f_i^{\sigma^*}(\Sigma^*)) \text{ in } \sigma \} \quad (26)$$

The identification of  $f_i^{\sigma^*}(\Sigma^*)$  with the empty centered situation  $\dagger^*$  if  $i$  is the index of an unembedded extensional verb (see **Ax3**) then captures the substitution-allowance of constructions like (2a) (in (33); see 19). The identification of  $f_i^{[f_j^{\sigma^*}(\Sigma^*)]}(\Sigma^*)$  with  $f_j^{\sigma^*}(\Sigma^*)$  if  $i$  is the index of an extensional and  $j$  the index of its embedding attitude verb (see **Ax4**) captures the substitution-resistance of constructions like (4a) (in (34)).

$$f_i^{\sigma^*}(\Sigma^*) = \dagger^* \text{ if } i \text{ is the index of an unembedded extensional verb} \quad ((\mathbf{Ax3}))$$

$$f_i^{[f_j^{\sigma^*}(\Sigma^*)]}(\Sigma^*) = f_j^{\sigma^*}(\Sigma^*) \text{ if } i \text{ and } j \text{ are the indices of an extensional and an attitude verb, respectively} \quad ((\mathbf{Ax4}))$$

$$\begin{aligned}
 & \llbracket \text{the reaction indicates [that sodium is a metal]} \rrbracket(\sigma_0^*) & (27) \\
 = & \{\sigma \mid \text{the reaction indicates}^i \{\sigma' \mid \text{Na is a metal in } \sigma' \& \text{ Na has all properties} \\
 & \quad \text{from } \llbracket \text{sodium} \rrbracket(f_i^{\sigma_0^*}(\Sigma^*)) \text{ in } \sigma'\} \text{ in } \sigma\} \\
 = & \{\sigma \mid \text{the reaction indicates} \{\sigma' \mid \text{Na is a metal in } \sigma' \& \text{ Na has all properties} \\
 & \quad \text{from } \llbracket \text{sodium} \rrbracket(\dagger^*) \text{ in } \sigma'\} \text{ in } \sigma\} \\
 = & \{\sigma \mid \text{the reaction indicates} \{\sigma' \mid \text{Na is a metal in } \sigma'\} \text{ in } \sigma\}
 \end{aligned}$$

$$\begin{aligned}
 & \llbracket \text{Len believes [that the reaction indicates [that sodium is a metal]]} \rrbracket(\sigma_0^*) & (28) \\
 = & \lambda p \lambda x \lambda \sigma^*. \{\sigma \mid x^\circ \text{ believes}^i p(f_i^{\sigma^*}(\Sigma^*)) \text{ in } \sigma\} \\
 & (\lambda \sigma^*. \{\sigma' \mid \text{the reaction indicates}^j \{\sigma'' \mid \text{Na is a metal in } \sigma'' \& \text{ Na has all} \\
 & \quad \text{properties from } \llbracket \text{sodium} \rrbracket(f_j^{\sigma^*}(\Sigma^*)) \text{ in } \sigma''\} \text{ in } \sigma'\})(\llbracket \text{Len} \rrbracket)(\sigma_0^*) \\
 = & \lambda \sigma^*. \{\sigma \mid \text{Len believes}^i [\lambda \sigma^*. \{\sigma' \mid \text{the reaction indicates}^j \{\sigma'' \mid \text{Na is a metal} \\
 & \quad \text{in } \sigma'' \& \text{ Na has all properties from } \llbracket \text{sodium} \rrbracket(f_j^{\sigma^*}(\Sigma^*)) \text{ in } \sigma''\} \\
 & \quad \text{in } \sigma'\}] (f_i^{\sigma^*}(\Sigma^*)) \text{ in } \sigma\}(\sigma_0^*) \\
 = & \{\sigma \mid \text{Len believes}^i \{\sigma' \mid \text{the reaction indicates}^j \{\sigma'' \mid \text{Na is a metal in } \sigma'' \& \\
 & \quad \text{Na has all properties from } \llbracket \text{sodium} \rrbracket(f_j^{[f_i^{\sigma^*}(\Sigma^*)]}(\Sigma^*)) \text{ in } \sigma''\} \text{ in } \sigma'\} \text{ in } \sigma\} \\
 = & \{\sigma \mid \text{Len believes}^i \{\sigma' \mid \text{the reaction indicates}^j \{\sigma'' \mid \text{Na is a metal in } \sigma'' \& \\
 & \quad \text{Na has all properties from } \llbracket \text{sodium} \rrbracket(f_i^{\sigma^*}(\Sigma^*)) \text{ in } \sigma''\} \text{ in } \sigma'\} \text{ in } \sigma\}
 \end{aligned}$$

The substitution-resistance of (4a) is then explained by the difference between  $\llbracket \text{sodium} \rrbracket(f_i^{\sigma_0^*}(\Sigma^*))$  and  $\llbracket \text{natrum} \rrbracket(f_i^{\sigma_0^*}(\Sigma^*))$ . As a result, Integrated Semantics solves all of the problems of two-dimensional semantics from Sect. 2.2.

## 6 Conclusion and Future Work

We have shown that Integrated Semantics resolves the tension between compositionality (or uniformity of interpretation) and pluralism about linguistic content: the semantics provides a uniform interpretation of extensional and attitude verbs that predicts the substitution behavior of constructions containing these verbs and that captures the agent-dependent interpretation of attitude reports.

We have restricted our considerations in this paper to the integrated contents of proper names (as representatives for referential DPs) and have limited the interpretation of verbs and VPs to an update of the attitude content of the verbs' DP-arguments by the verbs' truth-conditional content. However, as is illustrated in (29), the substitutivity of equivalent

CPs in attitude reports may also depend on the attitude content of other syntactic CP-constituents (here: on the content of the constituent nouns *groundhog* and *woodchuck*).

- (29) a. Eve believes [<sub>CP</sub>that Phil is a groundhog]. (T)  
      $\not\Rightarrow$  b. Eve believes [<sub>CP</sub>that Phil is a woodchuck]. (F)

Future work will extend the interpretation of verbs and VPs from Sect. 4 to a contextually determined interpretation that also respects the verbs' cognitive content, and will provide IS-interpretations of expressions from other syntactic categories.

**Acknowledgements** I wish to thank two anonymous referees for CoSt16 for their valuable comments and suggestions. The paper has profited from discussions with Mark Bowker, Robin Cooper, Makoto Kanazawa, Nikola Kompa, Manfred Krifka, Sebastian Löbner, Peter Pagin, Chris Tancredi, Markus Werning, Dietmar Zaefferer, Ed Zalta, and Ede Zimmermann. The research for this paper is supported by LMU Munich's institutional strategy LMUexcellent and by the German Research Foundation (Deutsche Forschungsgemeinschaft) (via Ede Zimmermann's grant ZI 683/13-1).

## References

- Barsalou, L. W. (1992). Frames, concepts, and conceptual Fields. In A. Lehrer, & E. F. Kittay (Eds.), *Frames, Fields, and Contrasts* (pp. 21–74). Erlbaum Associate.
- Carnap, R. (1988). *Meaning and Necessity*. University of Chicago Press.
- Chalmers, D. J. (2006). The foundations of two-dimensional semantics. In M. Garcia-Carpintero, & J. Macia (Eds.), *Two-Dimensional Semantics* (pp. 55–140). Oxford UP.
- Ciardelli, I., Roelofsen, F., & Theiler, N. (2017). Composing alternatives. *Linguistics and Philosophy*, 40(1), 1–36.
- Ciardelli, I., & Roelofsen, F. (2018). An inquisitive perspective on modals and quantifiers. *Annual Review of Linguistics*, 4, 129–149.
- Cresswell, M. J. (1985). *Structured Meanings*. MIT Press.
- Frege, G. (1892). Über Sinn und Bedeutung, *Gottlob Frege, Kleine Schriften*, Edition Classic Verlag Dr. Müller, 2006, 143–162.
- Haas-Spohn, U. (1995). *Versteckte Indexikalität und Subjektive Bedeutung*. Doctoral dissertation, Tübingen University.
- Heim, I. (1998). *Anaphora and semantic interpretation: A reinterpretation of Reinhart's approach*, MIT Working Papers in Linguistics, 25, pp. 205–246.
- Von Heusinger, K. (2013). The salience theory of definiteness. In *Perspectives on Linguistic Pragmatics* (pp. 349–374). Springer.
- Hintikka, J. (1969). Semantics for propositional attitudes. In J. W. Davis, D. J. Hockney, & W. K. Wilson (Eds.), *Philosophical Logic* (pp. 21–45). Springer.
- Israel, D. & Perry, J. (1996). Where monsters dwell. *Logic, Language and Computation*, 1, 303–316.
- Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 489–563). Oxford UP.
- Kripke, S. A. (1980). *Naming and Necessity*. Harvard UP.
- Lappin, S. (2015). Curry typing, polymorphism, and fine-grained intensionality . In S. Lappin, & C. Fox (Eds.), *Handbook of Contemporary Semantic Theory* (2nd ed., pp. 408–428). Wiley-Blackwell.
- Lerner, J. Y., & Zimmermann, T. E. (1991). *Eigennamen*, Semantik: Ein internationales Handbuch, A. v. Stechow, & D. Wunderlich, (Eds.), pp. 349–370.
- Lewis, D. (1970). General semantics. *Synthese*, 22(1–2), 18–67.

- Liefke, K., & Werning, M. (2018). Evidence for single-type semantics—an alternative to e/t-based dual-type semantics. *Journal of Semantics*, 35(4), 639–685.
- Montague, R. (1970). Universal grammar. *Theoria*, 36(3), 373–398.
- Moschovakis, Y. (2006). A logical calculus of meaning and synonymy. *Linguistics and Philosophy*, 29(1), 27–89.
- Muskens, R. (1995). *Meaning and Partiality*. CSLI Publications.
- Perry, J. (2012). *Reference and Reflexivity* (2nd edn.) CSLI Publications.
- Pollard, C. (2015). Agnostic hyperintensional semantics. *Synthese*, 192(3), 535–562.
- Potts, C. (2005). *The logic of conventional implicatures*. Oxford UP.
- Quine, W. V. O. (1953). *Reference and modality, From a Logical Point of View*, 1980. Harper and Row, 139–159.
- Schlenker, P. (2003). A plea for monsters. *Linguistics and Philosophy*, 26(1), 29–120.
- Von Stechow, A., Zimmermann, T. E. (2004). A problem for a compositional treatment of de re attitudes, Reference and Quantification, G. N. Carlson, & F. J. Pelletier (Eds.), CSLI, pp. 207–228.
- Stephenson, T. (2010). Vivid attitudes: centered situations in the semantics of ‘remember’ and ‘imagine’. *Proceedings of SALT*, 20, 147–160.
- Theiler, N., Roelofsen, F., & Aloni, M. (2018). A uniform semantics for declarative and interrogative complements. *Journal of Semantics*, 35(3), 409–466.
- Thomason, R. H. (1980). A model theory for the propositional attitudes. *Linguistics and Philosophy*, 4(1), 47–70.
- Westerståhl, D. (2012). Compositionality in Kaplan style semantics. In W. Hinzen, E. Machery, & M. Werning (Eds.), *The Oxford Handbook of Compositionality* (pp. 192–219). Oxford UP.
- Zalta, E. N. (1997). A classically-based theory of impossible worlds. *Notre Dame Journal of Formal Logic*, 38(4), 640–660.
- Zimmermann, T. E. (2012). Context dependence. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Handbook of Semantics* 3, de Gruyter, pp. 2360–2407.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Counting Possible Configurations



Manfred Krifka

**Abstract** It is often assumed that a requirement for counting objects is that they do not overlap. However, this condition can be violated. The paper deals, specifically, with counting objects that consist of parts, that is, with configurations. One example is *outfit* as a configuration of articles of clothing; notice that one article of clothing may be part of different outfits. The article develops an analysis of such configurational entities as individual concepts. It investigates the interaction of noun phrases based on such nouns with modal operators and in collective and cumulative interpretations.

**Keywords** Counting · Configurational objects · Individual concepts · Modal operators

## 1 Introduction

One of the conditions for counting is that the atoms of counting should not overlap (cf. e.g. Rothstein 2010; Landman 2016). The reason for this is obvious: In cases in which the atoms of a count noun are not settled, only a non-overlap condition will provide us with a counting function that yields a unique number. For example, when asked how many *body parts* a person has, it would be misleading to count the left

---

The analysis presented in this article was first presented as “Counting Configurations” at *Sinn und Bedeutung* 13 in 2009 in Stuttgart, and published in the proceedings Arndt Riester and Torgrim Solstad (eds.), pp. 309–324. University Stuttgart: SinSpeC; it is currently unavailable there. The present article is substantially extended. I thank the audiences at that conference, and at the conference *Cognitive Structures* in September 2016 at the University of Düsseldorf. In particular, thanks to Regine Eckardt, Ilaria Frana, Hans-Martin Gärtner, Andreas Haida, Stefan Hinterwimmer, Sophie Repp, and Magdalena Kaufmann for helpful comments, and to Sebastian Löbner especially for comments on the proper analysis of the main example of this article, *outfit*. Of course, for any problems of the theory and its exposition, I am solely responsible.

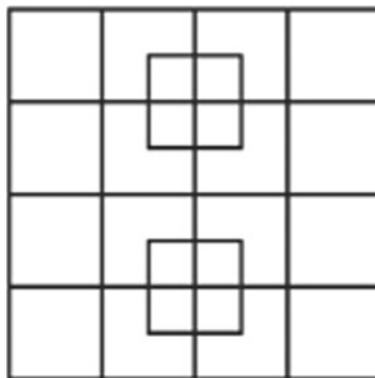
---

M. Krifka (✉)

Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), Humboldt-Universität Zu Berlin,  
Berlin, Germany

e-mail: [krifka@leibniz-zas.de](mailto:krifka@leibniz-zas.de)

arm, the left hand, and the five fingers of the left hand as distinct body parts, ending up with seven body parts on the left upper limb. Similarly, when asked how many *sequences of letters* there are in the set {abcde, hijkl, mnopq}, the answer 3 will be appropriate, but not 39, the number of sequences of two or more letters contained in these three maximal sequences.



However, there are contexts in which the disjointment requirement can be loosened up. There are riddles like *How many squares are in this figure?* which can perfectly well be answered by considering overlapping squares (in the picture to the right, this would result in 40 squares<sup>1</sup>). Counting overlapping entities may also be necessary in contexts that clearly are not riddles. For example, one study found that there are 5815 craters on the moon with a diameter  $\geq 20$  km, many of them overlapping.<sup>2</sup> Or we might want to know how many stories are actually contained in the Arabian Nights, which famously contains stories nested in stories—e.g., there is a story contained in a story contained in a story contained in a story contained in a story.<sup>3</sup> Even the counting of sequences might give rise to second thoughts, as the following entry in the discussion board for the board game *Sequence* shows<sup>4</sup>:

- (1) Just bought this game today and was playing with my young son. In the second game, he managed to construct a 6 in a row sequence. Now, this could be considered as being 2 overlapping 5 chip sequences. The rules are fairly sparse, but the strict definition is a sequence is "a connected series of five of the same colour marker chip in a straight line. If the definition was modified to "... five or more ..." then it would be clearer that you cannot overlap sequences in the same direction.

<sup>1</sup>Cf. <http://www.puzzlesandriddles.com/PerceptualPuzzle06.html>.

<sup>2</sup><https://cosmoquest.org/x/blog/2012/02/how-many-craters-are-on-the-moon/>.

<sup>3</sup>E.g. the e.g. the Tale of the Husband and the Parrot, see [https://en.wikipedia.org/wiki/List\\_of\\_stories\\_within\\_One\\_Thousands\\_and\\_One\\_Nights](https://en.wikipedia.org/wiki/List_of_stories_within_One_Thousands_and_One_Nights).

<sup>4</sup><https://boardgamegeek.com/thread/587189/why-dont-you-count-6-row-two-sets-5-sequences>.

In addition to entities that overlap in a given world, there are entities that arguably show overlap only in other worlds than the real world. Consider the following example<sup>5</sup>:

- (2) You have 3 shirts and 4 pairs of pants. *How many different outfits* can you make? [...] You get *twelve outfits*. Not counting if a dude makes an outfit without a shirt, or a crazy person without pants.

Assume we have three shirts  $s_1, s_2, s_3$  and four pairs of pants  $p_1, p_2, p_3, p_4$ , we can form twelve pairs of a shirt and a pair of pants, such as  $\langle s_1, p_1 \rangle, \langle s_2, p_2 \rangle, \langle s_2, p_1 \rangle$  and so on—twelve possible combinations altogether. Notice that the question here is not, *How many outfits are these?* The answer to that question would probably be *three*, if we count as an outfit a pair of a shirt and a pair of pants. That shirt  $s_1$  makes an outfit with  $p_1$  and also with  $p_2$  does not count, because we could not dress two persons at the same time with it. The question, rather, is *How many outfits can you make?*, where the modal *can* makes a crucial contribution. It requires that we look at different circumstances, where in some  $\langle s_1, p_1 \rangle$  makes an outfit, in others  $\langle s_1, p_2 \rangle$  makes an outfit.

Once one is aware of such cases, it is not difficult to find more, or to construct convincing examples at will<sup>6,7,8</sup>:

- (3) [Description of a tangram set.] With just seven simple pieces, you can make *dozens of amazing shapes*.
- (4) [Description of fischertechnik crane construction kit:] 100 Bauteile ermöglichen den Bau *dreier unterschiedlicher, einfacher Kräne*.  
‘With 100 construction parts enable one to build *three different, simple cranes*.’
- (5) [Description of Scrabble Word Builder:] We typed in the letters C, D, P, N, Y, E, A, and U and the Word Builder provided *dozens and dozens of words* that could be created with those letters.

<sup>5</sup>[answers.yahoo.com/question/index?qid=20080723031442AAcny3](http://answers.yahoo.com/question/index?qid=20080723031442AAcny3). The text continues: “Now let’s say you throw in three different pairs of socks...then you’d have 3 shirts times 4 pairs of pants times 3 pairs of socks for 36. It can get crazy the more options you throw in there.”

<sup>6</sup>[www.amazon.com/Think-Fun-4985-Tangram/dp/B000BXHP04](http://www.amazon.com/Think-Fun-4985-Tangram/dp/B000BXHP04).

<sup>7</sup>[spielwaren.lindex.de/Fischertechnik@Cranes@Fischertechnik@Basic.19673.WOB00000001.13](http://spielwaren.lindex.de/Fischertechnik@Cranes@Fischertechnik@Basic.19673.WOB00000001.13).

<sup>8</sup>[www.education-world.com/a\\_lesson/dailylp/dailylp/dailylp099.shtml](http://www.education-world.com/a_lesson/dailylp/dailylp/dailylp099.shtml).

Our main concern here is in the fact that even though these sentences talk about twelve outfits, dozens of tangram shapes, three cranes, and dozens and dozens of words, they do not imply that at any one possible world or point in time, dozens of shapes, twelve outfits, three cranes, or dozens of word tokens constructed with a set of eight scrabble pieces co-exist. Nevertheless, these sentences are true. The numeral constructions like *twelve outfits* appear to count things that exist across the different possible worlds or times referred to by the modal or temporal operators of the sentences. Notice that each of the sentences contains a modal marker, here underlined.

Perhaps this might not appear so remarkable for our examples if *tangram shape*, *crane*, or *word* refer to types (or kinds), which presumably have a more abstract way of existence anyway. But the examples can easily be read to refer to the concrete tangram pieces, construction parts, and scrabble letters in front of our eyes. And (2) does not lend itself to a type reading; the shirts and pants that are mixed and matched may well be unique.

## 2 The Problem with Configurations

I will assume that words like *outfit*, *tangram figure*, but also *crane* and *word*, apply to “configurations”. They refer to entities that consist of well-individuated parts that come together at certain worlds and times to form a certain configuration or to serve a purpose, but may be taken apart and be reconfigured at other indices. I take this to mean that words like *outfit* do not refer to regular individuals, type **e**, as this would not account for the fact that their parts are used to make up another individual at other indices.

To make our discussion more concrete, consider the following example, a simplified variant of (2).

- (6) It is possible to make four outfits with these two shirts and two pants.

We assume an interpretation format with explicit quantification over indices for worlds or times (including time intervals), and with entities that can be combined to form sum entities. I use  $i$ ,  $i'$  etc. as variables over indices (type **S**), and  $u$ ,  $u'$  etc. as variables over entities (type **e**), and I write  $u \sqcup u'$  for the sum (join) of  $u$  and  $u'$ , which is also of type **e** (cf. Link 1983 for the material join operation). Entities like outfits, tangram figures, cranes, and words are complex, as they typically consist of parts that are recognizable entities themselves. For example, a fischertechnik toy crane consists of various plastic pieces that are stuck together to resemble a crane, a tangram figure consists of the seven tangram pieces arranged in a way that iconically depicts another entity, an orthographic word consists of letters arranged in a linear sequence, and an outfit consists of articles of clothing that dress a person in a culturally acceptable way. The noun *outfit* comes with an additional complexity, as it is also a functional term (Löbner 2011); we speak of the *outfit of Mary* at a time  $t$  as the clothes that

Mary wore at  $t$ .<sup>9</sup> However, in examples like (2) it has a non-functional interpretation, and other nominals like *crane* and *tangram figure* that show the same configurational interpretation do not have a functional reading at all. In the non-functional reading of *outfit*, the person that is wearing the outfit is implicit, and the meaning of *outfit* could be given as follows:

- (7)  $\llbracket \text{outfit} \rrbracket = \lambda i \lambda u [u \text{ consists of articles of clothing worn by a person in } i, \text{ where the articles and their arrangement in } i \text{ satisfy the accepted dress codes in } i]$

According to this approach, the intension of *outfit* maps each index  $i$  to the set of entities  $u$  that consists of articles of clothes worn by a person at  $i$  in a way that follows the dress code at  $i$  (the latter provides for the facts like that a shirt and a pair of pants would not count as a complete outfit at an index with more formal standards).<sup>10</sup>

There is an implicit assumption in configurational objects like outfits that is important to be made explicit here: At any one index, an article of clothing can be used to dress only one person. We normally do not count two shirts and one pair of pants as two outfits, even if the pants are very large so that one slender person squeezes into each pant leg, and is additionally dressed by a shirt.

The numeral *four* can be represented in various ways. Let us assume here the standard Generalized Quantifier analysis, where  $P$  is a variable over properties, type *set*, and  $\#$  is a function that, when applied to a function of entities to truth values, type *et*, yields the number of entities that are mapped to the value 1, truth. In the Generalized Quantifier interpretation of numerals this is commonly assumed to be at least 4, in contrast to the quantifier *exactly four* (cf. Barwise and Cooper 1981).

- (8)  $\llbracket \text{four outfit} \rrbracket = \lambda i \lambda P [\# (\lambda u [\llbracket \text{outfit} \rrbracket(i)(u) \wedge P(i)(u)]) \geq 4]$

The predicate *make u with u'* is quite complex in its own right. For our purpose we understand it as follows: The agent selects parts of  $u'$  and creates an  $u$  out of these parts that did not exist immediately before. Following von Stechow (2001) on verbs of creation, I express this as in (9), where  $i' \angle i$  stands for “ $i'$  immediately precedes  $i$ ”, EXIST( $i$ ) identifies the entities that exist at the index  $i$ , and CONST( $i$ )( $u$ ) is the set of entities that  $u$  consists of at  $i$ .

- (9)  $\llbracket \text{make ... with ...} \rrbracket = \lambda i \lambda u \lambda u' \lambda u'' \exists i' [i' \angle i \wedge \neg \text{EXIST}(i')(u) \wedge u'' \text{ causes in } i': [\text{EXIST}(i)(u)] \wedge \forall u''' [u''' \in \text{CONST}(i)(u) \rightarrow u''' \sqsubseteq u']]$

<sup>9</sup>Thanks to Sebastian Löbner for pointing out the semantic complexities of *outfit*.

<sup>10</sup>In the functional reading, as present in expressions like *outfit of Mary*, the intension of *outfit* would consist in a function OUTFIT-OF =  $\lambda i \lambda u' \lambda u [u \text{ consists of articles of clothing worn by the person } u' \text{ in } i, \text{ provided that the satisfying the dress code in } i]$ . The sortal meaning we are interested in here can be derived by existential binding over the person argument  $u'$ , as  $\lambda i \lambda u \exists u' [\text{OUTFIT-OF}(i)(u') = u]$ , ‘outfit of a person’.

To illustrate, consider the following example, where *this* refers to the sum individual of two shirts  $s_1, s_2$  and two pants,  $p_1$  and  $p_2$ , rendered as  $s_1 \sqcup s_2 \sqcup p_1 \sqcup p_2$ .

- (10)  $\llbracket \text{John made an outfit with this} \rrbracket(i_0)$   
 $= \exists i[i < i_0 \wedge \exists u[\llbracket \text{outfit} \rrbracket(i)(u) \wedge \exists i'[i' \not\sim i \wedge \neg \text{EXIST}(i')(u) \wedge$   
 $\quad [\text{John causes in } i': [\text{EXIST}(i)(u)] \wedge$   
 $\quad \forall u''[u'' \in \text{CONST}(i)(u) \rightarrow u'' \sqsubseteq s_1 \sqcup s_2 \sqcup p_1 \sqcup p_2]]]]$

This means that at some time  $i$  in the past relative to  $i_0$ , John caused at an immediately preceding index  $i'$  that at  $i$  an entity is created that is an outfit at  $i$ , such that the things the outfit consists of are part of the two shirts and two pairs of pants referred to by *this*. We are not interested in a fine-grade analysis of causality here—this would state that there is some action on John's part at or before  $i'$  such that without that action the result, here that  $u$  exists, would not have been achieved (cf. Lewis 1973, based on the analysis of causality by David Hume). Also, we will not go into the CONST relation for now, but note here that it must allow for a newly created outfit to consist of parts that existed already before. Finally, it should be noted that we often understand (10) in a way that the person that wears the outfit at  $i$  is the agent, John, himself—but this need not be the case, e.g. if John is a fashion designer.

It is obvious that when  $s_1, s_2, p_1, p_2$  are the only articles of clothing, and any combination of a shirt and a pair of pants satisfies the dress code requirements for an outfit, the four combinations  $s_1 \sqcup p_1, s_1 \sqcup p_2, s_2 \sqcup p_1, s_2 \sqcup p_2$  are the only acceptable ones that can be used to create outfits. And as the same article of clothing cannot serve as part of two different outfits at the same index, sentence (11) cannot be true at any particular index  $i_0$ .

- (11)  $\llbracket \text{John made four outfits with this} \rrbracket(i_0)$   
 $= \exists i[i < i_0 \wedge [\#(\lambda u[\llbracket \text{outfits} \rrbracket(i)(u) \wedge \exists i'[i' \not\sim i \wedge \neg \text{EXIST}(i')(u) \wedge$   
 $\quad [\text{John causes in } i': [\text{EXIST}(i)(u)] \wedge$   
 $\quad \forall u''[u'' \in \text{CONST}(i)(u) \rightarrow u'' \sqsubseteq s_1 \sqcup s_2 \sqcup p_1 \sqcup p_2]]]] \geq 4]$

This is because (11) requires that four outfits exist at time  $i$ . We might think that the modality of the original example (2) helps. However, this is not the case. Consider the following simple interpretation of possibility:

- (12)  $\llbracket \text{it is possible} \rrbracket = \lambda i' \lambda p \exists i \in R(i')[p(i)]$

First, the modal may have wide scope with respect to the DP, resulting in the following interpretation at an index  $i_0$ .

- (13)  $\llbracket \llbracket \text{it is possible} \rrbracket [[\text{four outfits}]]_i [\text{to make } t_i \text{ with this}] \rrbracket(i_0)$   
 $= \lambda i[\llbracket \text{it is possible} \rrbracket(i)(\lambda i'[\llbracket \text{four outfits} \rrbracket(i')(\llbracket \text{to make with this} \rrbracket(i'))])](i_0)$   
 $= \exists i \in R(i_0)[\#(\lambda u[\llbracket \text{outfit} \rrbracket(i)(u) \wedge \exists i'[i' \not\sim i \wedge \neg \text{EXIST}(i')(u) \wedge$   
 $\quad \exists u''[u'' \in \text{CONST}(i)(u) \rightarrow u'' \sqsubseteq s_1 \sqcup s_2 \sqcup p_1 \sqcup p_2]]]] \geq 4]$

This states that there is some index  $i'$  accessible from  $i_0$  such that the cardinality of outfits made with the two shirts and two pairs of pants at  $i'$  is at least four. Clearly, this is not the intended reading: The sentence does not refer to a possible index in which, for example, a seamstress undoes the two shirts and two pants and makes four shirts and four pants out of them, thus creating four outfits in that world.

Second, the DP might have wide scope with respect to the modal. This results in the following interpretation:

$$\begin{aligned}
 (14) \quad & [[four\ outfits]]_i [it\ is\ possible\ [to\ make\ t_i\ with\ this]]](i_0) \\
 & = \lambda i [[four\ outfits]](i) (\lambda u [[it\ is\ possible]](i) (\lambda i' [[to\ make\ with\ this]](i'(u)))](i_0) \\
 & = \#(\lambda u [[outfit]](i_0)(u) \wedge \exists i \in R(i_0) \wedge \exists i' [i' \neq i \wedge \neg \text{EXIST}(i')(u) \wedge \\
 & \quad \exists u'' [u'' \text{ causes in } i': \text{EXIST}(i)(u)] \wedge \\
 & \quad \forall u''' [u''' \in \text{CONST}(i)(u) \rightarrow u''' \sqsubseteq s_1 \sqcup s_2 \sqcup p_1 \sqcup p_2])]) \geq 4
 \end{aligned}$$

This result is even worse because it states that there exist four outfits made with the two shirts and two pairs of pants at the index of interpretation  $i_0$  itself.

### 3 An Individual Concept Analysis

What went wrong? The problem is with the analysis of outfits as simple entities, type  $e$ . The representations in (11), (13) and (14) force us to assume that there are four outfits made of the two shirts and two pairs of pants at the same time. The solution I would like to propose is that outfits and their ilk are rather individual concepts, that is, functions from indices to entities, type  $se$ . Such functions may be partial, that is, they need not be defined for a particular index. In this case we say that the individual concept does not “exist” at that index, in the sense that it does not have a value. But it exists as a concept, as a function from indices to entities, and this concept may have properties, like being an outfit.

Individual concepts were used by Gupta (1980) to model the meaning of sentences like *National Airlines served two million passengers in 1975*. Gupta pointed out that this does not entail that National Airlines served two million persons, as one and the same person can perform the role of a passenger multiple times. Gupta’s solution—which analyzes passengers as individual concepts defined only for the time of a person’s flight—is problematic, as we find the same interpretation for sentences like *National Airlines served two million persons in 1975*, and persons, unlike passengers, are not individuated by flights (cf. Krifka 1990). But individual concepts appear to be well-suited for configurations.

To illustrate the individual concept analysis, take the four outfits one can make with the two shirts  $s_1, s_2$  and the two pairs of pants  $p_1, p_2$ . I make use of the notation introduced in Heim and Kratzer (1998) according to which an expression of the form  $\lambda v. \text{Restriction}[v]. [\text{Value}[v]]$  denotes the (possibly partial) function from entities of

the type of  $v$  that is only defined for arguments for which  $\text{Restriction}[v]$  holds; if defined, the function gives as value whatever is specified in  $\text{Value}[v]$ .

- $$(15) \begin{aligned} o_1 &= \lambda i. s_1 \text{ and } p_1 \text{ dress a person following cultural norms in } i. [s_1 \sqcup p_1] \\ o_2 &= \lambda i. s_1 \text{ and } p_2 \text{ dress a person following cultural norms in } i. [s_1 \sqcup p_2] \\ o_3 &= \lambda i. s_2 \text{ and } p_1 \text{ dress a person following cultural norms in } i. [s_2 \sqcup p_1] \\ o_4 &= \lambda i. s_2 \text{ and } p_2 \text{ dress a person following cultural norms in } i. [s_2 \sqcup p_2] \end{aligned}$$

For example,  $o_1$  is an individual concept that is only defined for indices  $i$  if the entities  $s_1$  and  $p_1$  dress a person following the dress code in  $i$ ; if defined,  $o_1$  maps to the sum entity consisting of the entities  $s_1$  and  $p_1$ . As one piece of clothing cannot be part of two outfits at any given index, the outfit concepts  $o_1$ ,  $o_2$  and  $o_3$  have non-overlapping domains and cannot exist at the same indices; only the outfits  $o_1$  and  $o_4$  (and the outfits  $o_2$  and  $o_3$ ) can co-exist, as they consist of non-overlapping parts.

It is clear what it means that an individual concept  $x$  exists at an index: It exists precisely at the indices in its domain. That is, if  $x$  is an individual concept, type  $\mathbf{s}(\mathbf{se})t$ , and  $\text{EXIST}$  is a predicate of individual concepts, type  $\mathbf{s}(\mathbf{se})t$ , then we have  $\text{EXIST}(i)(x) = 1$  iff  $i \in \text{DOM}(x)$ . For example, the concept  $o_1$  exists for all indices  $i$  for which  $o_1(i)$  is defined, that is, for which  $s_1 \sqcup p_1$  dresses a person following cultural norms in  $i$ . This means that  $o_1$  does not exist for all indices  $i$  at which  $s_1 \sqcup p_1$  does not dress a person, or else  $s_1 \sqcup p_1$  dresses a person but the cultural norms are so different that this does not follow the dress code. Consequently, the outfit  $o_1$  probably is of a rather punctuated or spotty nature: It may exist on May 1, then again on July 22, and on September 7, the times when  $o_1$  is actually used to dress a person, but not in the times in between.

Gupta analyzed common nouns as properties of individual concepts, type  $\mathbf{s}(\mathbf{se})t$ , and we will follow him in this respect. The common noun *outfit* applies to individual concepts like  $o_1$  in (15), and not to simple entities. I first give the extension of this common noun meaning at an index  $i_0$  in the set notation; it is of type  $(\mathbf{se})t$ .

- $$(16) \llbracket \text{outfit} \rrbracket(i_0) = \{\lambda i. u \text{ consists of articles of clothing worn by a person in } i, \text{ where the articles and their arrangement in } i \text{ satisfy the accepted dress code in } i_0 . [u] \mid u \in D_e\}$$

This is the set of all functions from indices  $i$  to entities  $u$  in the universe  $D_e$  whose parts are worn by a person in  $i$  and form an acceptable dress according to the standards of  $i_0$ . The condition about the parts of  $u$  are expressed by way of a restriction of this function. This accounts for the fact that there might be indices at which we do not consider the arrangement of a striped shirt and a checkered pair of pants a suitable outfit.

We can describe the intension of *outfit* as follows, in a first approximation:

- $$(17) \llbracket \text{outfit} \rrbracket = \lambda i' \lambda x \forall i \in \text{DEF}(x) [x(i) \text{ consists of articles of clothing worn by a person in } i, \text{ where the articles and their arrangement in } i \text{ satisfy the accepted dress code in } i']$$

Notice that it might happen that at a given index  $i_0$ , all the individual concepts in  $\llbracket \text{outfit} \rrbracket(i_0)$  are such that they are not defined for  $i_0$ , because none of them is worn in an acceptable way. Nevertheless,  $\llbracket \text{outfit} \rrbracket(i_0)$  is not empty in this case. To give a concrete example, assume a set of seven indices  $i_0, \dots, i_6$ , and assume that the four outfits mentioned in (15) are the following functions:

$$(18) \quad \begin{array}{ll} o_1 = [i_1 \rightarrow s_1 \sqcup p_1, i_2 \rightarrow s_1 \sqcup p_1] & \text{indices: } i_0 \quad i_1 \quad i_2 \quad i_3 \quad i_4 \quad i_5 \quad i_6 \\ o_2 = [i_4 \rightarrow s_1 \sqcup p_2, i_5 \rightarrow s_1 \sqcup p_2] & \text{outfits: } o_1 \quad o_1 \quad \quad \quad o_2 \quad o_2 \\ o_3 = [i_5 \rightarrow s_2 \sqcup p_1, i_6 \rightarrow s_2 \sqcup p_1] & \quad \quad \quad o_4 \quad o_4 \quad \quad \quad o_3 \quad o_3 \\ o_4 = [i_2 \rightarrow s_2 \sqcup p_2, i_3 \rightarrow s_2 \sqcup p_2] & \end{array}$$

Notice that  $o_1$  and  $o_4$  both are realized at  $i_2$ , and  $o_2$  and  $o_3$  both are realized at  $i_5$ , but that  $o_1$  and  $o_3$  as well as  $o_2$  and  $o_4$  do not co-exist. At  $i_0$  no outfit is realized at all. But the noun *outfit* denotes for all indices, including  $i_0$ , the set of all these individual concepts, if what qualifies as outfit is the same for all indices. The meaning of *outfit* is a constant property.

$$(19) \quad \llbracket \text{outfit} \rrbracket = \lambda i \in \{i_0, \dots, i_6\} \lambda x \in \{o_1, o_2, o_3, o_4\}$$

The meaning in (17) is not restrictive enough. In a situation like (18) it does not prevent us from calling, say, the function  $[i_1 \rightarrow s_1 \sqcup p_1]$  an outfit as well that is distinct from  $o_1$ , as it is only defined for the index  $i_1$ . Clearly, outfits are maximal with respect to indices, in the sense that for every index  $i$  at which  $s_1 \sqcup p_1$  is worn by a person, satisfying the dress code, this index belongs to the domain of the individual concept. Furthermore, in a situation like (18) we could not count an individual concept like  $[i_1 \rightarrow s_1 \sqcup p_1, i_4 \rightarrow s_1 \sqcup p_2]$  as an outfit, because it maps its indices to different articles of clothing. This violates the identity criteria that we normally assume for individual concepts, that they consist of the same entities, or the same substance.<sup>11</sup> A spelled-out version of (17) that includes these general conditions for cognitively relevant individual concepts would read as in (20), where the second line guarantees substance identity, and the third line maximality.

$$(20) \quad \llbracket \text{outfit}' \rrbracket = \lambda i' \lambda x \quad [(17) (i')(x) \wedge \begin{array}{l} \forall i, i'' \in \text{DEF}(x) [(17) (i')(x) \rightarrow x(i) = x(i'')] \wedge \text{same substance} \\ \neg \exists x' [(17) (i')(x') \wedge \text{DOM}(x') \subset \text{DOM}(x)] \end{array}] \quad \text{maximality}$$

The semantic type of *outfit*, a property that refers to individual concepts, would have to work with the expressions *outfit* combines with. For example, the predicate *wear* would have the following interpretation, where the object concept  $x$  is reduced to the value of  $x$  at the index of interpretation.

$$(21) \quad \llbracket \text{wear} \rrbracket = \lambda i \lambda x \lambda u [u \text{ is wearing } x(i) \text{ at } i]$$

---

<sup>11</sup>This is not quite true, as incremental changes are sometimes possible, cf. e.g. the example of the ship of Theseus, whose planks are replaced one by one over time, or living creatures that undergo metabolism, or entities like waves that consist in an ever-changing configuration. In all such examples there must be additional criteria of identity beyond material constituency.

Non-extensional predicates like *rise* or *change* are not reducible in this way (cf. Montague 1973).<sup>12</sup> This also applies to predicates of creation. The verb *make* states that an agent causes an individual concept to be realized at an index. For example, if John makes outfit  $o_1$  at index  $i$  then John causes that at  $i$ ,  $o_1$  becomes defined. This presupposes that during the making of  $i$ , the individual concept  $o_1$  was not defined (one cannot be making something that exists already) and involves some action by the agent on the parts that  $o_1$  refers to,  $s_1 \sqcup p_1$ , during the time before  $i$ . The essential parts of this is captured in the following interpretation.

$$(22) \llbracket \text{make} \dots \text{with} \dots \rrbracket = \lambda i \lambda x \lambda u \lambda u' \exists x' \exists i' [i' \not\sim i \wedge \neg i' \in \text{DOM}(x) \wedge u' \text{causes in } i':[i \in \text{DOM}(x)] \wedge x(i) \sqsubseteq u]$$

This states that at  $i'$  the individual concept  $x$  is not realized, but the agent  $u'$  causes that it is realized at the immediately following index  $i$ , where  $x(i)$  consists of parts of  $u$ .

The DP *four outfits* is interpreted as follows in the Generalized Quantifier analysis, where  $P$  is a variable for properties of individual concepts, type  $S(se)t$ .

$$(23) \llbracket \llbracket \text{DP four outfits} \rrbracket \rrbracket = \lambda i \lambda P [\#(\lambda x [\llbracket \text{outfit} \rrbracket(i)(x) \wedge P(i)(x)]) \geq 4]$$

We now can give an appropriate interpretation to our example. It states that there are four outfit concepts such that there are accessible indices at which these outfits are made. Notice that the predication is understood as distributive: For each of these individual concepts, there is an accessible index at which it can be made.

$$\begin{aligned} (24) \llbracket \llbracket \text{four outfits} \rrbracket \lambda t [\text{it is possible} [\text{to make } t \text{ with this}]] \rrbracket & (i_0) \\ &= \lambda i [\llbracket \llbracket \text{four outfits} \rrbracket \rrbracket(i)(\lambda x [\llbracket \text{it is possible} \rrbracket(i)(\lambda i' [\llbracket \text{to make with this} \rrbracket(i')(x)))])(i_0) \\ &= \llbracket \llbracket \text{four outfits} \rrbracket \rrbracket(i_0)(\lambda x [\llbracket \text{it is possible} \rrbracket(i_0)(\lambda i' [\llbracket \text{to make with this} \rrbracket(i')(x)))] \\ &= \#(\lambda x [\llbracket \text{outfits} \rrbracket(i_0)(x) \wedge \llbracket \text{it is possible} \rrbracket(i_0)(\lambda i' [\llbracket \text{to make with this} \rrbracket(i')(x)))] \geq 4 \\ &= \#(\lambda x [x \in \{o_1, o_2, o_3, o_4\} \wedge \\ &\quad \exists i' \in R(i_0) \exists u' \exists i'' [i'' \not\sim i' \wedge \neg i'' \in \text{DOM}(x) \wedge \\ &\quad [u' \text{ causes in } i'':[i'' \in \text{DOM}(x)] \wedge x(i') \sqsubseteq s_1 \sqcup s_2 \sqcup p_1 \sqcup p_2,]]]) \geq 4 \end{aligned}$$

This is true iff for each of the four individual concepts  $x$  there is an index  $i'$  accessible from  $i_0$  such that  $x$  is realized by someone at  $i'$ , and  $x(i')$  is part of the two shirts and two pants. It is crucial that this does not entail that there is an index at which all four individual concepts are realized simultaneously. In particular, (23) is compatible with a situation in which only two outfits can be realized at a time.

It should be pointed out that there is also a consistent interpretation for the following example, if the quantifier scopes over the past tense operator:

---

<sup>12</sup>The verb *change* can be used for outfits in its functional sense, as in *Mary changed her outfit*. Let us represent the functional reading OUTFIT-OFF as  $\lambda i \lambda u \lambda x [\llbracket \text{outfit}' \rrbracket(i)(x) \wedge u \text{ is wearing } x(i) \text{ at } i]$  (alternatively, we can start with a functional reading and derive the sortal reading, as in footnote 10). Then *Mary changes her outfit* is true at  $i$  iff there is an  $i'$  shortly before  $i$ , and an  $i''$  shortly after, such that OUTFIT-OFF( $i'$ )(Mary)  $\neq$  OUTFIT-OFF( $i''$ )(Mary).

$$\begin{aligned}
 (25) \quad & \llbracket \text{John made four outfits with this} \rrbracket(i_0) \\
 & = \llbracket \llbracket \text{four outfits} \rrbracket \lambda t \text{ PAST}[\text{John makes with this}] \rrbracket(i_0) \\
 & = \#(\lambda x[x \in \{o_1, o_2, o_3, o_4\} \wedge \\
 & \quad \exists i'[i' < i_0 \wedge \exists i''[i'' \angle i' \wedge \neg i'' \in \text{DOM}(x) \wedge \\
 & \quad [\text{John causes in } i'':[i' \in \text{DOM}(x) \wedge x(i') \sqsubseteq s_1 \sqcup s_2 \sqcup p_1 \sqcup p_2]]]) \geq 4
 \end{aligned}$$

The sentence can be true at a given index, as the individual concepts may come into existence at different times; notice that the existential quantifier  $\exists i' \dots$  has scope under the quantifier *four outfits*.

In this section I have proposed a semantic interpretation of sentences like (2) that stays close to the standard Generalized Quantifier analysis of sentences with numerically modified nouns like *four outfits*. The only substantial change is that the noun *outfit* does not refer to ordinary individuals, but to individual concepts. In the next section we will argue that the individual concept analysis should be generalized; it should apply to entities such as shirts and pairs of pants as well.

## 4 Generalizing the Individual Concept Analysis

### 4.1 Is Everything an Individual Concept?

There are good reasons to apply the individual concept analysis to other individuals than to just configurational individuals, like outfits. Take, for example, Ludwig Wittgenstein; he can be represented by an individual concept that maps all indices  $i$  at which Wittgenstein exists to Wittgenstein—in our world, these are all indices from April 26, 1889 to April 29, 1951. In contrast to the domains of configurations that fade in and out of existence, this is a convex set of indices: If  $i$  and  $i'$  are indices of the same possible history that are in this set, and if  $i''$  is an index of the same possible history that is temporally in-between  $i$  and  $i'$ , then  $i''$  is in this set as well.<sup>13</sup> As another example, take role concepts like *the tallest woman*, or *the Pope*. In contrast to configurations, such concepts may refer to different entities for different indices. As a third example, take individual concepts like the denotation of *the gifted mathematician that John claims to be* (cf. Grosu and Krifka 2008). Such expressions denote individual concepts that refer to the same entity, but are restricted to those indices that are compatible with John's claims. The individual concept analysis also affords for analyses of concepts like a wave (of water), which has a convex set of indices but maps these indices to ever-changing water entities.

If regular individuals are also based on individual concepts, then this also should hold for pants and shirts. After all, they certainly are created, and destroyed. As

---

<sup>13</sup>The individual concept view opens a new way to deal with modified names, like *(the) young Mozart* (cf. Paul 1994), as a subconcept; the term refers to the same entity as *Mozart* but is only defined for those indices at which Mozart was young.

individual concepts, they differ from outfits insofar as they have a convex domain: Whenever it holds that  $i, i' \in \text{DOM}(x)$ , then for all  $i''$  that are temporally in between,  $i < i'' < i'$ , it also holds that  $i'' \in \text{DOM}(x)$ . But what, then, do individual concepts map their indices to? We might think of the substance or matter they consist of (this corresponds to the  $h$  homomorphism in Link (1983) that maps objects to matter). Hence, the shirt  $s_1$  would be also of type  $\mathbf{se}$ , a function from indices to the matter the shirt consists of, provided that this matter forms a shirt at these indices. Moreover, for concrete objects like shirts we have to assume additional conditions, namely that the matter is more or less the same between indices, allowing for occasional small changes like replacing a button in the case of a shirt, or metabolic exchanges of matter in the case of living creatures.

The outfit  $o_1$  consisting of the shirt  $s_1$  and the pair of pants  $p_1$ , which are analyzed as concrete object concepts themselves, can then be defined as follows:

$$(26) \quad o_1 = \lambda i. s_1 \text{ and } p_1 \text{ dress a person following cultural norms in } i. [s_1(i) \sqcup p_1(i)]$$

That is,  $o_1$  maps every index  $i$  for which it is defined to the same stuff as the join of the stuff of  $s_1$  and  $p_1$  at  $i$ . In general, we would have the following interpretation of *outfit* as a property of individual concepts; an appropriate maximalization as in (20) would have to be added.

$$(27) \quad \llbracket \text{outfit} \rrbracket = \lambda i' \lambda x \forall i \in \text{DEF}(x) [x \text{ consists of articles of clothing worn by a person in } i, \text{ where the articles and their arrangement in } i \text{ satisfy the accepted dress code in } i']$$

The only difference to (17) is that  $x$ , not  $x(i)$ , is required to consist of articles of clothing. That is, for each outfit  $x$  there must be articles of clothing  $x_1, x_2, \dots, x_n$  such that  $x = x_1 \sqcup x_2 \sqcup \dots \sqcup x_n$ . The material join operation for individual concepts is defined as follows:

$$(28) \quad x \sqcup y = \lambda i[x(i) \sqcup y(i)]$$

This is an individual concept that is defined for all indices for which  $x$  and  $y$  are defined, and maps these indices to the sum of  $x$  and  $y$ . This leads to the following definition, where  $\sqcup P$  is the join of all individuals in the set  $P$ .

$$(29) \quad \llbracket \text{outfit} \rrbracket = \lambda i' \lambda x \exists P \forall i \in \text{DEF}(x) [\forall y \in P [y \text{ is an article of clothing in } i'] \wedge x = \sqcup P \wedge \exists z[\text{person}(i)(z) \wedge \text{dressed-with}(i)(z)(x) \wedge \text{satisfies-dress-code}(i')(z)(x)]].$$

This says that whenever  $x$  is an outfit, then it applies to the same matter as the sum of some set  $P$  of articles of clothing. The sum of the matter of these articles of clothing is the same as the matter of the outfit, but the articles of clothing may be defined for a larger, and typically convex, domain. Even though (28) does not require this

literally, we can think of each outfit  $x$  being associated with a unique set of articles of clothing  $P$ .

## 4.2 Coercion to Constituting Parts

Commenting on an earlier version of this article, Sebastian Löbner suggested an analysis in which entities like outfits are regular entities, type  $e$ , instead of individual concepts. The idea is that any combination of entities that can form an outfit is in the extension of *outfit*; in our example, these are the entities  $s_1 \sqcup p_1$ ,  $s_1 \sqcup p_2$ ,  $s_2 \sqcup p_1$ ,  $s_2 \sqcup p_2$ . This suggests the following interpretation, where outfits are of type  $e$ :

- (30)  $\llbracket \text{outfit} \rrbracket(i_0) = \lambda u \exists i \in R(i_0)[u \text{ is worn by a person in } i \text{ in a way that satisfies the dress codes in } i]$

Note that under this analysis *outfit* still has an intensional component (an entity  $s \sqcup p$  of type  $e$  is an outfit iff in some possible world  $i$ , a person wears  $i$ , and this satisfies the dress code in  $i$ ). But the intensionality is not hard-wired in the notion of objects itself, which remain of type  $e$ . They are not lifted to individual concepts,  $se$ .

A problem of this analysis is that it does not motivate the use of creation verbs like *make*, *bauen* ‘build’ and *create* in examples (2)—(5). If the outfit  $o_1$  is identical to the sum of entities  $s_1 \sqcup p_1$ , what does it mean to *make an outfit*? It would perhaps refer to the tailor’s sewing of the shirt and the pair of pants, but not to the person that combines this shirt and this pair of pants to wear them together, as suggested in example (2). For this reason, the individual concept analysis, even though it is more complex, appears appropriate.

On the other hand, the interpretation (29) would allow for a straightforward analysis of examples like (30) that are problematic for the interpretations in (17) or (26).

- (31) There is an outfit in the wardrobe.

When we understand outfits as entities that are defined only when someone wears them, then (30) could not be true, except in the peculiar case of a person sitting in the wardrobe and wearing an outfit.

I assume that individual concepts with spotted realizations like outfits can be coerced into individual concepts with a more permanent interpretation, and it is these coerced concepts that are involved in sentences like (30). If (30) refers to  $o_1$ , which consists of the concrete individual concepts  $s_1$  and  $p_1$ , then  $o_1$  can be coerced into the individual concept  $s_1 \sqcup p_1$  as defined in (27). Let us call this coercion “grounding”. Then (30) states that this sum concept is in the wardrobe.

Grounding in general can be interpreted as the following function:

(32) Grounding (coercion to parts)

For any individual concept  $x$ ,  
 if there are cognitively salient individual concepts  $x_1, \dots, x_n$   
 such that  $\forall i \in \text{DOM}(x)[x(i) = x_1(i) \sqcup \dots \sqcup x_n(i)]$ ,  
 then  $g(x) = x_1 \sqcup \dots \sqcup x_n$

Let us assume that  $s_1$  and  $p_1$ , a shirt and a pair of pants, are modeled by individual concepts as suggested in example (25). The outfit  $o_1$  would have the following grounded version:

(33)  $g(o_1) = \lambda i[s_1(i) \sqcup p_1(i)]$

This is the individual concept that has the same domain as  $o_1$  and always refers to the sum of the shirt  $s_1$  and the pants  $p_1$ . As  $s_1$  and  $p_1$  have convex domains, so has  $g(o_1)$ , the grounded version of  $x_1$ . In particular  $g(o_1)$  does also exist at indices  $i$  at which no-one is wearing  $s_1$  and  $p_1$  as an outfit; in our small model (18),  $g(o_1)$  exists at all indices from  $i_0$  to  $i_6$ . Consequently,  $g(o_1)$  can have the property of being in the wardrobe at an index like  $i_0$  at which  $o_1$  does not have any realizations.

It should be noted that, as  $g$  is a function,  $g(x)$  presupposes that there is a unique cognitively most salient way to analyze  $x$  as consisting of concrete objects  $x_1, \dots, x_n$ . These are the elements in the set  $P$  in (28). For an outfit, these are the articles of clothing, but not their parts, like the buttons, buckles and the pieces of cloth that they consist of. As they may have existed before the shirt, and may exist after, their sum may lead to an individual concept with a longer duration. If a unique decomposition could not be guaranteed, we would have to model  $g$  not as a function, but as a relation that maps  $x$  to different decompositions.

When predicates like *be in the wardrobe* are applied to individual concepts, then we can assume coercion by the grounding operation triggered by the meaning of the predicate. This is because such predicates can be reduced to the matter that an individual concept realizes at an index, which requires coercion to a more permanent entity<sup>14</sup>:

(34)  $\llbracket \text{an outfit is in the wardrobe} \rrbracket(i_0)$

$$\begin{aligned} &= \exists x[\llbracket \text{outfit} \rrbracket(i_0)(x) \wedge \llbracket \text{in the wardrobe} \rrbracket(i_0)(g(x))] \\ &= \exists x[\llbracket \text{outfit} \rrbracket(i_0)(x) \wedge g(x)(i_0) \text{ is in the wardrobe at } i_0] \end{aligned}$$

This is true at  $i_0$  iff  $x$  is an outfit, as before, and the things  $x$  consists of—the shirt  $s_1$  and the pair of pants  $p_1$ —are in the wardrobe.

Grounded individual concepts can also explain the use of creation verbs to refer to the entities an individual concept consist of, as in *the tailor made an outfit*. In this case, the object is coerced to its grounded interpretation, due to the knowledge of what tailors typically create.

---

<sup>14</sup>This reduction from individual concepts to stages is similar to the reduction from individuals to stages for stage-level predicates in Carlson (1977).

### 4.3 Joining and Counting Individual Concepts

Grounded individual concepts can be counted, as the examples (34) and (35) show.

(35) There are 100 outfits in this wardrobe.

(36) The supply department ordered 100 new outfits for the employees.

Obviously, the regular individual concept analysis of *outfit* in (17) does not work, as outfits are not worn by anyone when they are in the wardrobe or when they are ordered. But the entity analysis (29) and the grounded individual concept analysis (31) also are problematic. They would make our examples true in case 10 shirts and 10 pants that can be randomly combined to outfits are in the wardrobe, or are ordered, because they can be configured to 100 outfits. This is illustrated for the grounded individual concept analysis in (37).

(37)  $\#\{x \mid [\![\text{outfit}]\!](i_0)(x) \& [\![\text{in the wardrobe}]\!](i_0)(g(x))\} = 100$   
 if  $g(x) = \lambda i[s(i) \sqcup p(i)]$ ,  $s \in \{s_1, \dots, s_{10}\}$ ,  $p \in \{p_1, \dots, p_{10}\}$ , and both conjuncts are true

The problem here is that configurational individual concepts like outfits defy the usual property of additivity under the current interpretation. Additivity would tell us that if  $x$  is one outfit, and  $y$  is another outfit, then  $x$  and  $y$  together are two outfits. However, as we have seen, we might end up with four outfits. This is because outfits, other than ordinary individuals like shirts and pants, can overlap. The generalized quantifier strategy of representing numbers inherent in (36) cannot rule out such counting of overlapping objects. A theory that fares better is the one proposed in Krifka (1995), according to which count nouns are measure functions that can be applied to sum entities, and specify the number of the things they are applied to. We need something like count noun variants of nominal predicates (marked here by \*) that map individual concepts to numbers, and that follow the rule of additivity:

(38) a.  $[\![\text{outfit}^*]\!](i)(x) = 1$  if  $x$  consists of one outfit, i.e.  $[\![\text{outfit}]\!](i)(x)$   
 b.  $[\![\text{outfit}^*]\!](i)(x) = n \wedge [\![\text{outfit}^*]\!](i)(x') = n' \wedge x, x'$  do not overlap at  $i$   
 $\rightarrow [\![\text{outfit}^*]\!](i)(x \oplus x') = n + n'$

Here,  $x \oplus x'$  stands for the sum of the individual concepts  $x$  and  $x'$ . Notice that (37)(a) and (b) happen to be the same standardization and generalization operations proposed in Krifka (1990) for event-related quantification. Arguably, they belong to the general conceptual tool kit for constructing measure functions in language.

But what does the sum of two individual concepts actually mean? This would need detailed elaboration; I can give here just the basic construction steps. We have to assume that the domain of individual concepts, type **se**, has a sum structure. Let AIC the set of all atomic individual concepts; this is the set of individual concepts as considered so far. The set of sum individual concepts SIC then is defined as the smallest set such that (a)  $\text{AIC} \subseteq \text{SIC}$ , and whenever  $x, x' \in \text{SIC}$ , then also  $x \oplus x' \in$

SIC. Here,  $\oplus$  is a join operation that is idempotent, commutative, and associative. We understand it in such a way that the resulting set SIC is homomorphic to the power set of all individual concepts, with atomic individual concepts  $x$  represented by singletons,  $\{x\}$ , and sum individual concepts like  $x \oplus x'$  represented by set union,  $\{x\} \cup \{x'\} = \{x, x'\}$ .

Sum individual concepts are still functions from indices to entities. In particular, a sum individual concept maps an index to the sum of the parts when they are defined for that index. That is, we require that  $[x \oplus x'](i) = x(i) \sqcup x'(i)$ , if  $x$  and  $x'$  are defined for  $i$ ;  $[x \oplus x'](i) = x(i)$ , if only  $x$  is defined for  $i$ , and  $[x \oplus x'](i) = x'(i)$ , if only  $x'$  is defined for  $i$ . Notice that different sum individual concepts can have the same functional value. For example, take  $w$  to be the Wittgenstein individual concept from 1889 to 1951,  $w_y$  the individual concept of the younger Wittgenstein defined from 1889 to 1921, and  $w_o$  the concept of the older Wittgenstein defined from 1922 to 1951, then  $w$  and  $w_y \oplus w_o$  are different sum individuals, but have the same value.

For (37) we still have to define what it means that two individual concepts  $x, x'$  overlap at an index  $i$ ; this is the case if there is an entity that is a part of  $g(x)$  at  $i$  and a part of  $g(x')$  at  $i$ , that is, if there is an  $u$  such that  $u \sqsubseteq g(x)(i)$  and  $u \sqsubseteq g(x')(i)$ .

The truth conditions of an example like (34), here simplified, can be rendered as follows:

$$(39) \quad [\![\text{there are two outfits in the wardrobe}]\!](i_0) = \\ \exists x[\![\text{outfit}^*]\!](i)(x) = 2 \wedge [\![\text{in the wardrobe}]\!](i_0)(g(x))]$$

The sentence is intuitively true under our assumption that the outfit  $o_1$  made of  $s_1$  and  $p_1$  and the outfit  $o_4$  made of  $s_2$  and  $p_2$  are in the wardrobe. According to (37)(a), it holds that  $[\![\text{outfit}^*]\!](i_0)(o_1) = 1$  and  $[\![\text{outfit}^*]\!](i_0)(o_4) = 1$ , and as  $o_1$  and  $o_2$  do not overlap, we have  $[\![\text{outfit}^*]\!](i_0)(o_1 \oplus o_2) = 2$ . Even if the outfit  $o_2$ , made of  $s_1$  and  $p_2$ , and outfit  $o_3$ , made of  $s_2$  and  $p_1$ , are also in the wardrobe, they could not be counted because they overlap with  $o_1$  and  $o_4$ . We of course could also sum up  $o_2$  and  $o_3$  instead, which would yield the same result.

It appears that sentences like *There are four outfits in the wardrobe* are felt to be ambiguous by some speakers, and can be considered true in one reading in which there are only two shirts and two pants in the wardrobe. This second reading can be generated by another construction of measure functions that differs from (37) by requiring that in the additivity clause, it is sufficient that  $x \neq x'$ , that is,  $x$  and  $x'$  may in fact overlap. Such weakened cases of additivity that allow for overlap are also relevant in cases like counting craters on the moon.

#### 4.4 Collective and Cumulative Interpretations

Having sum individual concepts also enables the interpretation of collective interpretations as in the following case:

- (40) Two (of the) outfits are rather similar to each other.

This is a predication on a sum individual concept, which is true iff the atomic parts stay in a similarity relation to each other (the strong interpretation of reciprocals; for weaker interpretations see Dalrymple et al. (1998) and subsequent literature on the “strongest meaning hypothesis”). Here,  $\leq_a$  is the atomic part relation on sum individual concepts.

$$(41) \exists x[\llbracket outfit^*\rrbracket(x_0)(x) = 2 \wedge \forall x' \forall x''[x' \leq_a x \wedge x'' \leq_a x \wedge x' \neq x'' \rightarrow \llbracket similar \rrbracket(i_0)(x', x'')]]$$

The interpretation of expressions like *two outfits* proposed here is also possible for the non-collective examples we started out with, provided that we assume that verbal predicates, when applied to sets of individual concepts, distribute over their elements. Instead of (23) we can entertain the following analysis:

$$(42) \llbracket [it \text{ is possible to make four outfits with this}] \rrbracket(i_0) \\ = \exists x[\llbracket outfit^*\rrbracket(i_0)(x) = 4 \wedge x = s_1 \oplus s_2 \oplus p_1 \oplus p_2 \wedge \\ \forall x'[x' \leq_a x \rightarrow \exists i' \in R(i_0) \exists y \exists i''[i'' \angle i' \wedge -i'' \in \text{DOM}(x') \wedge \\ [y \text{ causes in } i'': [i'' \in \text{DOM}(x)]]]]]$$

This states that there are four outfits  $x$  that consist of the two shirts and two pants, and that it is possible for each atomic part  $x'$  of  $x$  that some agent  $y$  brings it about to be realized.

Sum individual concepts are also relevant for cumulative interpretations (Scha 1981). Assume that a kindergarten owns a construction set with which all kinds of vehicles can be constructed, but only one at a time (there are only four wheels in the construction set).

- (43) Dozens of children have built hundreds of vehicles with this construction set.

Such interpretations have been explained as a consequence of the cumulativity of verbal predicates (cf. Krifka 1989; Sternefeld 1998). That is, transitive predicates like *build* are interpreted such that if  $x$  builds  $y$  and  $x'$  builds  $y'$ , then  $x \oplus x'$  builds  $y \oplus y'$ . This interpretation is triggered by Sternefeld’s operator  $**$ , here adapted as in (43), where  $R$  stands for the verbal predicate, type  $s(se)(se)t$ , and  $\leq$  is the part relation for sum individual concepts.

$$(44) **R = \lambda i \lambda x \lambda y [\forall x' \leq x \exists y' \leq y [R(i)(x')(y')]] \wedge \forall y' \leq y \exists x' \leq x [R(i)(x)(y)]]$$

This allows for the following representation of (42) at an index  $i_0$ , where “ $\gg 24$ ” and “ $\gg 200$ ” state that a number is in the range of dozens and hundreds, respectively.

- (45)  $\exists x \exists y [\llbracket \text{child}^* \rrbracket(i_0)(x) >> 24 \wedge \llbracket \text{vehicle}^* \rrbracket(i_0)(y) >> 200 \wedge \llbracket \text{built} \rrbracket(i_0)(y)(x)],$   
 where  $\llbracket \text{built} \rrbracket(i_0)(y)(x) =$   
 $\forall x' \leq x \exists y' \leq y \exists i' \exists i'' [i' < i_0 \wedge i'' \leq i' \wedge \neg i'' \in \text{DOM}(x') \wedge$   
 $[y \text{ causes in } i'': [i'' \in \text{DOM}(x)]']]] \wedge$   
 $\forall y' \leq y \exists x' \leq x \exists i' \exists i'' [i' < i_0 \wedge i'' \leq i' \wedge \neg i'' \in \text{DOM}(x') \wedge$   
 $[y \text{ causes in } i'': [i'' \in \text{DOM}(x)]']]$

This states that there is a sum individual concept  $x$  that are dozens of children and a sum individual concept  $y$  that are hundreds of vehicles, and that each part of the children built some part of the vehicles, and each part of the vehicles were built by some of the children. This renders the cumulative reading of (43) in an adequate way.

## 5 The Property Analysis

In this paper I have argued for individual concepts in our conceptual representation, and in particular, for the ability to count individual concepts. There is a proposal on a related topic, “Counting Concepts” by Condoravdi et al. (2001), which analyzes examples like the following in a way that looks similar to what we have proposed for configurations.

- (46) The mayor prevented three strikes.

*Prevent* is analyzed as an intensional predicate, like *seek*, which Condoravdi et al. (2001) interpret, following Zimmermann (1992), as having a property argument:

- (47)  $\llbracket \text{The mayor prevented a strike} \rrbracket(i_0)$   
 $= \exists i < i_0 [\llbracket \text{prevent} \rrbracket(i)(\llbracket \text{strike} \rrbracket)(\llbracket \text{the mayor} \rrbracket)]$   
 $= \exists i < i_0 [\llbracket \text{prevent} \rrbracket(i)(\lambda i' \lambda u [u \text{ is a strike in } i'])(\text{the mayor})]$

This captures the reading in which no reference to a specific strike is intended. The object DP, *a strike*, denotes a property of entities.

There is also a specific reading: There was a threat for a strike that was about to form, and the mayor prevented that strike from happening. The normal solution for specific reading, giving the noun phrase wide scope (cf. (3)), does not work. It entails the existence of a strike  $u$ —but this is exactly what the next conjunct says was prevented.

- (48)  $\exists i < i_0 \exists u [\llbracket \text{strike} \rrbracket(i)(u) \wedge \llbracket \text{prevent} \rrbracket(i)(\lambda i' \lambda v [u=v])(\text{the mayor})]$

Condoravdi et al. propose a solution for the specific interpretation using “subconcepts” (that is, subproperties). No strict definition is given, but we certainly should assume that a superconcept applies to all indices and individuals a subconcept applies

to. The specific reading of *the mayor prevented a strike* can be given as follows, where  $\subseteq_{sc}$  is the subconcept relation.

$$(49) \exists P [P \subseteq_{sc} [\text{strike}] \wedge \exists i[i < i_0 \wedge [\text{prevent}](i)(P)(m)]]$$

For the interpretation of *three strikes*, Condoravdi et al. (2001) discuss various options, settling on a generalized quantifier analysis:

$$(50) [[\text{the mayor prevented three strikes}]](i_0) \\ = \#(\lambda P[P \subseteq_{sc} [\text{strike}]] \wedge \exists i[i < i_0 \wedge [\text{prevent}](i)(P)(\text{the mayor})]) \geq 3$$

But for this to work, the notion of subconcept must be properly restricted. One entity may fall under different subconcepts of *strike*, e.g. it might be a strike of the railroad workers and at the same time (as railroad workers are public workers) a strike of the public workers. Obviously, the subconcepts that we count should not be such that one is included in the other. Hence Condoravdi et al. propose to restrict counting to minimal subconcepts, that is, to “maximally specific instantiated concepts”.

The use of minimal subconcepts suggests that we actually better work with individual concepts, because then we get minimality for free, as individual concepts can apply to maximally one entity. Hence it seems natural to propose the individual concept analysis to examples of this type as well. The natural reading of (45) is that what the mayor prevented was that three specific strike threats led each to a full-blown strike. In each world at which these strikes would have been realized, there would have been exactly one realization.

$$(51) [[\text{the mayor prevented three strikes}]](i_0) \\ = \exists x[[\text{strike}^*]](i_0)(x) = 3 \wedge \forall x' \leq_a x \exists i' < i_0 [\text{the mayor prevented } x' \text{ at } i']]$$

This says that the three strikes consists of three individual concepts  $x$  that are strikes, and that for each  $x'$  there exists an index  $i'$  in the past of the actual time  $i_0$  such that the mayor prevented at  $i'$  the strike  $x'$  from happening. Where *prevent* denotes a rather involved concept; it means that the subject referent (here: the mayor) caused the object referent (here:  $x'$ ) not to be realized, which means in turn that, if the mayor would not have acted then  $x'$  would exist for all normal continuation of  $i'$ .

But there is still an issue of identity to be considered: For example, assume that an announced strike is declared illegal, and the workers plan another strike with similar goals and methods to circumvent the court ruling, but this strike is declared illegal as well. In which sense can we say that two strikes were declared illegal? This depends on rather specific criteria. Formal semantics can only provide the general format of the objects of lexical semantics.

## 6 Conclusion

In this paper I have discussed the meaning of sentences that contain reference to what I called “configurational” objects, as denoted by such terms as *outfit* or *tangram figure*, or even *crane* and *word*. Configurational objects consist of parts that can be reconfigured, and exist only at those indices in which they stand in the appropriate configuration. I have argued that configurational objects can fruitfully be analyzed as individual concepts, functions from indices to entities. I have developed ways how such concepts can be counted in count-noun constructions like *four outfits*. I then argued that more regular entities like shirts should also be represented by individual concepts, albeit with more stable temporal properties, and I have shown that there are contexts in which a configurational object like *outfit* can actually be coerced to the object it consists of.

The general direction of this paper points towards a theoretical framework in which the objects referred to in language, and consequently, the objects of our cognition, should be seen as individual concepts. The notion of an object contains the ability to identify the same object over different indices, and this is precisely achieved by individual concepts. Some objects are temporally convex in the sense that they have a continuous existence from an initial time to a final time (such as shirts and pants), others have a more spotted existence (such as outfits). There are various other examples of objects with apparently extraordinary identity criteria, such as waves. Whether this view is suitable, or even sustainable, cannot be answered in this short paper. At least I hope to have shown that it provides us with ways to give truth conditions to sentences that count configurational objects.

## References

- Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4, 159–219.
- Carlson, G. N. (1977). A unified analysis of the English bare plural. *Linguistics and Philosophy*, 1, 413–456.
- Condoravdi, C., Crouch, R., & van den Berg, M. (2001). Counting concepts. In R. van Rooij & M. Stokhof (Eds.), *Proceedings of the thirteenth Amsterdam colloquium* (pp. 67–72).
- Dalrymple, M., et al. (1998). Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy*, 21, 159–210.
- Grosu, A., & Krifka, M. (2007). ‘The gifted Mathematician that you claim to be’: Equational intensional reconstruction relatives. *Linguistics and Philosophy*, 30, 445–485.
- Gupta, A. (1980). *The logic of common nouns: An investigation in quantified modal logic*. New Haven: Yale University Press.
- Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. London: Blackwell.
- Krifka, M. (1989). Nominal reference, temporal constitution and quantification in event semantics. In R. Bartsch, J. van Benthem, & P. van Emde Boas (Eds.), *Semantics and contextual expressions* (pp. 75–115). Dordrecht: Foris.
- Krifka, M. (1990). Four thousand ships passed through the lock: Objectinduced measure functions on events. *Linguistics and Philosophy*, 13, 487–520.

- Krifka, M. (1995). Common nouns: A contrastive analysis of Chinese and English. In: G. N. Carlson & F. J. Pelletier (Eds.), *The generic book* (pp. 398–411). Chicago, London: The University of Chicago Press.
- Landman, F. (2016). Iceberg semantics for count nouns and mass nouns: Classifiers, measures and portions. *Baltic International Yearbook of Cognition, Logic and Communication*, 11.
- Lewis, D. (1973). Causation. *The Journal of Philosophy*, 70, 556–567.
- Link, G. (1983). The logical analysis of plurals and mass terms: A latticetheoretical approach. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, use and the interpretation of language* (pp. 303–323). Berlin, New York: Walter de Gruyter.
- Löbner, S. (2011). Concept types and determination. *Journal of Semantics*, 28, 279–333.
- Montague, R. (1973). The proper treatment of quantification in ordinary English. In K. J. J. Hintikka, J. M. E. Moravcsik, & P. Suppes (Eds.), *Approaches to natural language* (pp. 221–242). Reidel: Dordrecht.
- Paul, M. (1994). Young Mozart and the Joking Woody Allen. Proper names, individuals and parts. In *SALT IV* (pp. 268–281). Cornell: CLC Publications.
- Rothstein, S. (2010). Counting and the mass-count distinction. *Journal of Semantics*, 27, 343–397.
- Scha, R. J. H. (1981). Distributive, collective, and cumulative quantification. In J. A. G. Groenendijk, T. M. V. Janssen, & M. B. J. Stokhof (Eds.), *Formal methods in the study of language* (Vol. 135, pp. 483–512). Amsterdam: Mathematical Centre Tracts.
- Sternfeld, W. (1998). Reciprocity and cumulative predication. *Natural Language Semantics*, 6, 303–337.
- von Stechow, A. (2001). Temporally opaque arguments in verbs of creation. In C. Cechetto, G. Chierchia, & M.T. Guasti (Eds.), *Semantic interfaces: Reference, anaphora, aspect* (pp. 278–391). Stanford: CSLI Publications.
- Zimmermann, T. E. (1992). On the proper treatment of opacity in certain verbs. *Natural Language Semantics*, 1, 149–180.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Structure and Ontology in Nonlocal Readings of Adjectives



Marcin Morzycki

**Abstract** In certain uses, adjectives appear to make the semantic contribution normally associated with adverbs. These readings are often thought to be a peripheral phenomenon, restricted to one corner of the grammar and just a handful of lexical items. I'll argue that it's actually considerably more general than is often recognized, and that it admits two fundamentally different modes of explanation: in terms of the syntactic machinery that undergirds these structures and in terms of the ontology of the objects manipulated by its semantics. Both modes of explanation have been suggested for some of the puzzles in this domain, and I'll argue both are necessary. With respect to adjectives including *average* and *occasional*, the key insight is that their lexical semantics is fundamentally about kinds. But to arrive at a more general theory of adverbial readings, it is also necessary to further articulate the compositional semantics. In this spirit, I'll argue that these adjectives actually have the semantic type of quantificational determiners like *every*. If this way of thinking about adverbial readings is on the right track, it instantiates a means by which these two distinct modes of explanation—and the distinct aspects of cognition they may ultimately be associated with—both play a crucial role in bringing about the apparently aberrant behavior of this class of adjectives.

**Keywords** Adjectives · Nonlocal readings · Average · Occasional · Kinds · Natural language metaphysics

## 1 Introduction

It is, of course, not news that the way language organizes the world may tell us something about how the mind does so. Nor is it news that that perhaps the best window into how language organizes the world is how language works: what words mean and how grammars manipulate those meanings. This is the project that Emmon Bach

---

M. Morzycki (✉)  
University of British Columbia, Vancouver, Canada  
e-mail: [marcin.morzycki@ubc.ca](mailto:marcin.morzycki@ubc.ca)

memorably dubbed ‘natural language metaphysics’ (Bach 1986, 1989) or ‘natural language ontology’. Importantly, it’s a project that’s worthwhile even if—perhaps especially if—it should fail to coincide with metaphysics proper, because our theory of natural language metaphysics is a repository of linguistic analysis. If we’re doing it right, its structure explains the structure of language. The structure of the world is another matter entirely, best left to others.

There is an important trade-off in this domain, however, that I’d like to use to frame this paper. Structures in natural language ontology can serve to explain linguistic phenomena, and when they do, they may lighten the explanatory burden on other components of linguistic theory, including the syntax and semantics. Conversely, introducing complexity in the syntax or semantics can make possible a simpler ontology.

It may help to sketch an example of what I have in mind. It’s entirely independent of the one I’ll focus on primarily in this paper. It concerns polar antonyms of adjectives, such as *tall* and *short*. No one would defend the view that they are unrelated, of course, so only the question is where to install a theory of their difference. One possibility is ontological. Height is measured in abstract representations of measurement, *degrees*, which include things like ‘6 feet’. The set of degrees that measure height (as opposed to e.g. weight) tell us the dimension along which a given measurement exists, but they don’t actually tell us whether we’re measuring how tall someone is or how short. They tell us that the dimension is spatial extent, but not whether the scale is tallness or shortness. To know that, we must know at least one more thing: the ordering imposed on those degrees. ‘6 feet’ is a greater degree of tallness than ‘5 feet’, but a lesser degree of shortness. On such a theory, advocated in Kennedy (1997, 2001) and elsewhere, the key to the relation between *tall* and *short* is that they measure on scales that impose opposite orderings on degrees along the same dimension. Their denotations therefore need not reflect any direct connection between the adjectives beyond specifying which scale they use, because the connection is between the scales, not between adjectives themselves.

The alternative is to suppose that the relation between *tall* and *short* is a matter of grammar, not (primarily) ontology, and that they use precisely the same scale after all. One might suppose, with Heim (2006, 2008) and Büring (2007), that *short* involves a special kind of negation, present in the syntactic tree but not normally pronounced as a separate morpheme. *Short*, on this view, is actually a way of pronouncing ‘little tall’ or ‘untall’. There are a variety of arguments to be made for this more complicated syntax, and with it in place, the ontology needn’t provide an independent analysis of the connection between the two antonyms because the richer syntax already does.

It’s not the case, of course, that any analysis of any arbitrary phenomenon can be said to be primarily grammatical or ontological. In the context of this volume, it’s especially worth noting that an approach that involves decomposition into features might occupy an intermediate position with respect to this distinction: the decomposition is in some respects like decomposing *short* into ‘little tall’, but of course the decomposition needn’t be implemented directly in the syntax in this way, and there are interesting discussions to be had about the relationship between decomposing word meanings and decomposing the underlying concepts themselves. The

former seems still a robustly grammatical enterprise; the latter considerably more an ontological one.

All that said, at some point there's a danger of putting more weight on this distinction than it can bear. Its purpose here is chiefly just to situate another empirical puzzle for which a balance has to be struck between grammatical and ontological explanation: adjectives like *average*. The first thing to notice is the curious effect they often have on the referent of the nominal in which they occur:

- (1) The average American has 2.3 children.

This sentence, Carlson and Pelletier (2002) point out, is doubly mysterious. What sort of entity is 'the average American'? Certainly, on its most natural reading, it doesn't refer to some particular American who is especially typical of Americans. Second, what sort of entity is '2.3 children'? If *the average American* referred to a particular American—say, one named Steve—it would suggest, alarmingly, that Steve has only a fraction of one of his children. That's not what the sentence means, at least ordinarily. Nor, indeed, is it possible to straightforwardly disentangle the strangeness of the first nominal from the strangeness of the second. Even if we avoid the reading under which (1) involves direct reference to Steve, it still fails to communicate that it is typical for Americans to have fractional children.

On its face, it would seem that to avoid such morally outlandish outcomes, we must embrace a metaphysically outlandish one. We must accept that there are such things as 'average Americans' in the model underlying the semantics, and indeed perhaps in some extended sense such things as '2.3 children'. I don't think we should dismiss this possibility too readily. For one thing, as Bach would remind us, our judgment in these matters must be guided by language, not a priori notions about what sorts of objects populate the actual world. That's the difference between natural language metaphysics and metaphysics proper. Indeed, this metaphysical direction is precisely the one in which Carlson and Pelletier head. For this reason, Hornstein (1984) was ultimately mistaken in saying that 'no one wishes to claim that there are objects that are average men in any meaningful sense'. Yet, he argued, nominals like *the average American* act no different from more referentially pedestrian ones. He concluded that this was an argument against the enterprise of formal semantics itself.

My aim here will be more modest. Kennedy and Stanley (2009) observed that sentences such as (1) can be analyzed as a special case of a more general phenomenon: readings of adjectives in which the adjective is interpreted as though it were an adverbial. This requires a more complex syntax, but that more complex syntax is a low price to pay for the metaphysical benefit. It frees us from having to posit any spookily abstract and therefore implausible entities in the ontology. I'll argue, building on Morzycki (2016b), these adverbial readings are in fact part of a considerably more general pattern of readings available to a far wider range of adjectives than generally recognized. I'll argue that these readings actually fall into three classes, and that this leads us to an analysis distinct in important respects from Kennedy and Stanley but that, as they argued, places the explanatory burden on the syntax and compositional semantics rather than the ontology.

In Sect. 1, following largely the argument in Morzycki (2016b), I'll present the case that what I'll call nonlocal readings of adjectives (following Schwarz 2006 et seq.) are far more general than is typically recognized, and that they fall into three distinct classes. In Sect. 2, I'll review some ways these problems have been approached in the past, highlighting the interplay between grammatical and ontological explanation. In Sect. 3, I'll propose a strategy for approaching these facts that I hope may eventually scale up to the larger empirical picture and that has components of both kinds of explanation. In particular, I'll combine elements of syntactic assumptions that have widely been made with a new ingredient in the compositional semantics: the idea that adjectives with external readings have determiner-like meanings, and as a consequence have the complex grammar associated with determiners. I'll sketch this idea in general terms for *average* in particular, relating it to Gehrke and McNally (2010, 2015)'s crucial insight that adjectives like *occasional* involve reference to kinds. Finally, in Sect. 4, I'll very briefly return to the larger issues with which we began: the analytical balance between structure in the syntax and semantics and structure in the ontology.

## 2 Nonlocal Readings of Adjectives

### 2.1 *On ‘Occasional’*

Let's begin with the classic example of a nonlocal reading of an adjective, which is *occasional* (Bolinger 1967; Stump 1981; Larson 1999; Zimmermann 2003; Schäfer 2007; Gehrke and McNally 2010, 2015; DeVries 2010). It's the best-studied such case, and this will serve as a useful background against which to consider *average*. The standard sentence is (2):

- (2) An occasional sailor strolled by.  
 a. INTERNAL: ‘Someone who sails occasionally strolled by.’  
 b. EXTERNAL: ‘Occasionally, a sailor strolled by.’

It has what's called an internal and an external reading. The internal reading is interesting in a number of respects, but from our current perspective, it's the external reading that is most immediately relevant. On this reading, the adjective makes a semantic contribution that is, to all appearances, completely divorced from the nominal in which it finds itself. The sailors that strolled by are sailors simpliciter. There is no question about the frequency of their sailing. But the situation is more puzzling still. On the external reading, the sentence means more or less the same thing as (3), where the definite determiner replaces the indefinite:

- (3) The occasional sailor strolled by.

Yet the meaning is essentially the same (but see Gehrke and McNally 2015 for detailed discussion). Indeed, some adjectives of this class (*odd* and *rare*) have the external reading *only* with *the*.<sup>1</sup> Setting apart a subtle change of flavor, the external reading also occurs with *your* and in the bare plural:

- (4) a. Your occasional sailor strolled by.  
 b. Occasional sailors strolled by.

So there are three mysteries so far: an ambiguity, unexpectedly wide scope, and unexpected interpretations of the determiner.

There are more still. Another is that, on the external reading, the adjective must occupy the leftmost position in the structure of the nominal:

- (5) The angry occasional sailor strolled by.  
 a. INTERNAL: ‘Someone angry who sails occasionally strolled by.’  
 b. #EXTERNAL: ‘Occasionally, an angry sailor strolled by.’

Indeed, the range of determiners with which *occasional* is possible on the external reading is relatively limited:

- (6)  $\left\{ \begin{array}{l} \text{Every} \\ \text{Some} \\ \text{Several} \\ \text{Many} \\ \text{Most} \end{array} \right\}$  occasional sailor(s) strolled by.

- a. INTERNAL: ‘*D* person/people who sail(s) occasionally strolled by.’  
 b. #EXTERNAL: ‘Occasionally, *D* sailor(s) strolled by.’

Yet another idiosyncrasy of the external reading is that it renders the adjective unable to coordinate with ordinary adjectives:

- (7) The occasional and angry sailor strolled by.  
 a. INTERNAL: ‘Someone angry who sails occasionally strolled by.’  
 b. #EXTERNAL: ‘Occasionally, an angry sailor strolled by.’

Another still: on this reading, the adjective becomes incompatible with degree words such as *very* or the comparative<sup>2</sup>:

---

<sup>1</sup>Berit Gehrke (p.c.) points out that this fact doesn’t follow from what will be proposed here—but then, I don’t really have an analysis to offer here of the *occasional* class more generally. That said, this fact is precisely what one might expect if, as Larson (1999) has argued, syntactic incorporation into a determiner gives rise to some lexical idiosyncrasy here. See Sect. 5 for more.

<sup>2</sup>For some speakers, even the internal reading is missing. Others can get an external reading marginally with *very*, but not with *more*.

- (8) The very occasional sailor strolled by.
  - a. INTERNAL: ‘Someone who sails very occasionally strolled by.’
  - b. #EXTERNAL: ‘Very occasionally, a sailor strolled by.’
  
- (9) The more occasional sailor strolled by.
  - a. INTERNAL: ‘Someone who sails more occasionally strolled by.’
  - b. #EXTERNAL: ‘More occasionally, a sailor strolled by.’

## 2.2 *Returning to ‘Average’*

Having noted the crucial features of *occasional*, let’s return to *average* with them in mind. First, there was ambiguity. As Carlson and Pelletier (2002), Kennedy and Stanley (2009) among others noted, there is an ambiguity with *average* too:

- (10) An average American has 2 children.
  - a. INTERNAL: ‘An American, who is typical, has 2 children.’
  - b. EXTERNAL: ‘On average, an American has 2 children.’

For the internal reading to be available without counterpragmatically ghastly background assumptions, we must change our earlier sentence to *2 children*. On this reading, the claim is that there is an American somewhere that is typical and that he has two children. There is another reading of *average* that also occurs in (11) (Sebastian Löbner, p.c.), which is also internal, or in any case fails to be external:

- (11) He’s so average.

The external reading is the one with which we are now familiar from *occasional*. It’s worth noting that it paraphrases naturally with an adverbial, *on average*, which is analogous to how *occasional* morphed into *occasionally*.

Here we encounter a set of properties that elegantly mirror those of *occasional*. There are unexpected interpretations of the determiner. Switching to the definite determiner leaves us, on the external reading, with apparently the same interpretation, and *your* is not much different:

- (12) {The  
Your} average American has 2 children.
  - a. INTERNAL: ‘{The  
Your} American that’s a typical one has 2 children.’
  - b. EXTERNAL: ‘On average, an American has 2 children.’

Again, on the external reading, other determiners don’t seem to work:

- (13) # $\left\{ \begin{array}{l} \text{Every} \\ \text{Most} \\ \text{Some} \\ \text{Several} \\ \text{Two} \end{array} \right\}$  average American(s) has/have 2.3 children.

And again, on the external reading *average* has to be leftmost among the adjectives in its nominal:

- (14) a. An average irritable American has 2.3 children.  
 b. #?An irritable average American has 2.3 children.

It is also unable to coordinate with other adjectives on the external reading:

- (15) #An irritable and average American has 2.3 children.

It is incompatible with degree modifiers on this reading:

- (16) #A very average American has 2.3 children.

So, once again, the same mysterious patterns manifested themselves as with *occasional*. At a minimum, this supports the connection between the two that Kennedy and Stanley (2009) posited—perhaps indeed more robustly than they intended. But the pattern is more widespread still.

## 2.3 *Wrong*

Before considering the bigger picture, it will be necessary to lay out a few more examples of the general phenomenon. A version of the now-familiar pattern emerges once again with *wrong* (Haik 1985; Schmitt 2000; Schwarz 2006, 2019). It too has an internal/external ambiguity, though perceiving it is slightly trickier. Suppose Floyd is a spy who is required to provide his interlocutor with false information and deprive her of true information. If he succeeds in this, (17) is true on the internal reading, on which the information provided was incorrect:

- (17) Floyd gave the wrong answer.  
 a. INTERNAL: ‘Floyd gave an answer that was incorrect.’  
 b. EXTERNAL: ‘Floyd gave an answer that it was wrong of him to give.’

On the external reading, (17) is false, because Floyd answered as he is supposed to. On the other hand, if Floyd slips up at some point and accidentally answers a question truthfully, the situation is flipped: (17) is still true, but only on the external reading: he provided information that he isn’t supposed to provide, namely, true information. Something similar happens in (18):

- (18) Floyd killed the wrong person.
- INTERNAL: 'Floyd killed a person that was wrong (perhaps prone to error or wrong in general).'
  - EXTERNAL: 'Floyd killed a person that it was wrong of him to kill.'

Again, the internal reading in (18) is more easily discerned with some context. Consider a dystopian game show in which participants are executed for answering a quiz question incorrectly. Floyd is the executioner. If he killed the contestant that answered incorrectly, (18a) is true only on the internal reading. ('Clyde was wrong, so I killed him,' he might explain.) If Floyd accidentally killed a contestant that provided the correct answer, (18b) would be true only on the external reading.

There is again an odd fact about the interpretation of the determiner: *the* is interpreted as an indefinite. In (17), there need not have been only one wrong answer, and in (18), there need not have been only one person who must not be killed. The picture is slightly different, though. *Your* is impossible here except on its usual possessive reading, irrelevant here:

- (19) a. ?Floyd gave your wrong answer.  
 b. ?Floyd killed your wrong person.

Strangely, it's not just that the definite determiner is interpreted as an indefinite, but it's the principal way to say this. The indefinite would be unusual on the external reading:

- (20) a. Floyd gave a wrong answer.  
 b. Floyd killed a wrong person.

It's not actually fully clear what reading these receive. For me, an external reading is possible, but only when there is a desire to communicate that there are multiple answers that shouldn't be given and people that shouldn't be killed.

Apart from that quirk, again we encounter restrictions on the choice of determiner on the external reading:

- (21) #Floyd opened  $\left\{ \begin{array}{l} \text{every} \\ \text{most} \\ \text{some} \\ \text{several} \\ \text{two} \end{array} \right\}$  wrong envelope(s).

As before, inherently quantificational determiners fail.

The requirement that the nonlocal adjective be structurally higher than other adjectives again emerges:

- (22) a. Floyd opened the wrong brown envelope.  
 b. #Floyd opened the brown wrong envelope.

So does the ban on coordination:

- (23) #Floyd opened the wrong and brown envelope.

And so does the ban on degree modification:

- (24) #Floyd opened the very wrong envelope.

So a rather large class of adjectives that includes *wrong*, *average*, *typical*, *occasional* and a number of its synonyms seems to manifest quite a number of common properties.

## 2.4 ‘Whole’ and ‘Entire’

The parallels continue with *whole* and *entire*, though there will be an important twist. As before, there is an ambiguity (Moltmann 1997, 2005; Morzycki 2002), which I’ll assume is a special case of the internal/external ambiguity:

- (25) A whole ship was submerged.
- a. INTERNAL: ‘A complete, structurally intact ship was submerged.’
  - b. EXTERNAL: ‘A ship was wholly submerged.’
- (26) The whole apple is terrible.
- a. INTERNAL: ‘The complete, structurally intact apple, the one with no bites taken out of it, is terrible.’
  - b. EXTERNAL: ‘All parts of the apple are terrible.’

The internal reading is actually the unusual one in these cases, and may take a moment to perceive. It’s what could be expressed more or less unambiguously with *complete*—indeed, I suspect that it’s precisely the existence of this unambiguous alternative that accounts (on broadly Gricean grounds) for the unnaturalness of the internal reading.

As before, there are restrictions on the determiner, but they take a different form. First, *a*, *the*, and *your* retain their usual meanings, and don’t become interchangeable. Second, strong quantifiers are still incompatible with the external reading, but weak ones are perfectly compatible with it (I will now indulge in the habit of marking sentences with a # when they are impossible on the external reading)<sup>3</sup>:

- (27) a.  $\left\{ \begin{array}{l} \# \text{Every} \\ \# \text{Most} \\ \text{Many} \\ \text{Several} \\ \text{Two} \end{array} \right\}$  whole ship(s)  $\left\{ \begin{array}{l} \text{was} \\ \text{were} \end{array} \right\}$  submerged.

---

<sup>3</sup>Sebastian Löbner (p.c.) points out that one might explain the ill-formed examples in (27) because one nominal can’t express two different quantifications (Löbner 2000), which would accord with the grammaticality of adverbial *entirely* in e.g. *Every ship was entirely white*.

- b.  $\left\{ \begin{array}{l} \#Every \\ \#Most \\ Many \\ Several \\ Two \end{array} \right\}$  whole apple(s)  $\left\{ \begin{array}{l} \text{is} \\ \text{are} \end{array} \right\}$  terrible.

The other, now increasingly familiar restrictions reemerge in their customary form. The external reading is only possible when the nonlocal adjective occurs high:

- (28) a. A whole enormous ship was submerged.  
 b. An enormous whole ship was submerged. *(internal only)*

It's incompatible with coordination:

- (29) A whole and enormous ship was submerged. *(internal only)*

And it's incompatible with degree modification:

- (30) An entirely whole ship was submerged. *(internal only)*

## 2.5 Epistemic Adjectives

Abusch and Rooth (1997) observed a proposition-modifying interpretation of what they called ‘epistemic adjectives’ that now won’t come as a shock. These adjectives include *unknown*, *undisclosed*, *unspecified*, and *unexpected*. They can receive a wide-scope reading:

- (31) Solange is staying at an unknown hotel. *(Abusch and Rooth 1997)*  
 a. INTERNAL: ‘Solange is staying at a hotel no one has heard of.’  
 b. EXTERNAL: ‘Solange is staying at a hotel and it is not known which hotel she is staying at.’

The external reading systematically supports concealed-question paraphrases. For many years in the early 2000s, (32) was a kind of running joke in American political discourse, and it’s actually very hard to make sense of its internal reading:

- (32) Dick Cheney is hiding at an undisclosed location.

The external reading is that Dick Cheney is hiding at a location and it has not been disclosed, for his safety, what location that is. On its internal reading, perhaps it would have to be the very fact that it is a location that is not disclosed.

At this stage, we will encounter the same empirical refrain, and the reader can presumably sing along. On the external reading, there are again restrictions on the determiner. Although *the* and *a* seem to behave normally, strong inherently quantificational determiners remain impossible:

- (33) Solange stayed at  $\left\{ \begin{array}{l} \# \text{every} \\ \# \text{most} \\ \text{some} \\ \text{several} \\ \text{two} \end{array} \right\}$  unknown hotel(s).

As for *whole*, weak determiners are compatible with external readings.

The restrictions on the structural position of the adjective in the DP remain the same. The external reading is, as we have come to expect, possible only when the adjective is high<sup>4</sup>:

- (34) a. Solange stayed at an unknown horrible hotel.  
 b. Solange stayed at a horrible unknown hotel. *(internal only)*

The external reading is unavailable when the adjective occurs in a coordinate structure:

- (35) #Solange stayed at a horrible and unknown hotel. *(internal only)*

It's incompatible with degree modification:

- (36) #Solange stayed at a very unknown hotel. *(internal only)*

## 2.6 Same and Different

Other adjectives fall under broadly the same rubric. Among the best-studied of these are *same* and *different* (Nunberg 1984; Heim 1985; Carlson 1987; Keenan 1992; Moltmann 1992; Beck 2000; Lasersohn 2000; Majewski 2002; Alrenga 2006, 2007a, b; Barker 2007; Brasoveanu 2011). The facts in this domain are complicated in ways that muddy the waters considerably, and the terminology is different and confusing, but for our purposes the important point is that there is an ambiguity.

The main terminological confound is that the internal reading involves an anaphoric dependency on preceding discourse. This is in an important sense 'external', but it is not external in the *relevant* sense of seeming to require the adjective to access to the semantic content of the clause outside the nominal itself. This is clearer when considering the readings:

---

<sup>4</sup> Sebastian Löbner suggests a mode of explanation of this fact: the concealed question-style semantics reveals these nominals denote individual concepts, which is incompatible with the sort of run-of-the-mill extensional intersective adjectival modification attempted in (34) and (35). Perhaps that strategy could help with the quantificational facts in (33) as well.

- (37) Floyd and Clyde read the same book.
- INTERNAL (ANAPHORIC): ‘Floyd and Clyde read a book that is the same as the one previously mentioned.’
  - EXTERNAL: ‘Floyd and Clyde read a book in common.’
- (38) Floyd and Clyde read a different book.
- INTERNAL (ANAPHORIC): ‘Floyd and Clyde read a book that is the different from the one previously mentioned.’
  - EXTERNAL: ‘The book Floyd read was not the same book as the one Clyde read.’

I won’t rehearse the full song-and-dance yet again, in part because it presents, in this instance, complications that go considerably beyond the scope of this paper. Suffice it to say that on the external reading, *same* and *different* impose restrictions on the determiner with which they combine:

- (39) \*Floyd and Clyde read { every  
          most  
          some  
          several  
          two } same book(s).

On this reading *same* and *different* are subject to the now familiar structural position requirement:

- (40) a. Floyd and Clyde read the same good book.  
       b. \*Floyd and Clyde read the good same book.

## 2.7 *Modal Superlatives: ‘Possible’ and Its Kin*

There is another important class of nonlocal readings of adjectives, which I will mostly set aside. These involve *possible*, *conceivable*, and the like ('modal superlatives'; Bolinger 1967; Larson 2000; Schwarz 2005; Cinque 2010; Romero 2013; Leffel 2014):

- (41) They interviewed every possible candidate.
- EXTERNAL: ‘They interviewed every candidate that it was possible to interview.’
  - INTERNAL: ‘They interviewed every person who was possibly a candidate.’

There are important distinctions between these cases and the ones we’ve examined so far, but for the moment I will note only the similarity: again, there is an ambiguity between an internal and external reading.

## 2.8 *Miscellaneous Obscurities and Novelties*

Without further discussion, I'll note a few examples of nonlocal readings that are either obscure or, to my knowledge, novel:

- (42) The inevitable counterexample arose.  
‘Inevitably, a counterexample arose.’
- (43) He spooned a moody forkful. (P.G. Wodehouse; Hall 1973)  
‘Moodily, he spooned a forkful.’
- (44) An unlikely chiropractor discovered the solution.  
‘A chiropractor discovered the solution and it was unlikely that that chiropractor (or a chiropractor?) would do so.’
- (45) Clyde asked a random linguist.  
‘Clyde asked a linguist randomly.’
- (46) Floyd received an unfortunate grade.  
‘Floyd received a grade such that it was unfortunate to receive it.’

One shouldn't read too much into these without careful examination, of course, but they collectively suggest that more external readings lurk just over our analytical horizon.

## 3 Three Classes of Nonlocal Readings

This paper is not a linguistic curio cabinet. We've established, I hope, that there are patterns in this domain. That's not to say that there aren't genuine mysteries here. It's just that the phenomena at issue are mysterious *in parallel ways*. The next stage is to systematize the patterns more robustly so we can move toward an analysis.

There are, I will argue, three distinct classes of nonlocal adjectives. The first class I will set aside here. It includes the aforementioned ‘modal superlatives’ like *possible*. They differ from the others most strikingly in which determiners are involved in the external readings. In these cases, universal quantifiers license the external reading, not inhibit it:

- (47) We interviewed  $\left\{ \begin{array}{l} \text{every} \\ \#\text{the} \\ \#\text{a} \\ \#\text{no} \\ \#\text{three} \end{array} \right\}$  possible candidate.

Superlatives and *only* also license it:

- (48) We interviewed  $\left\{ \begin{array}{l} \text{the only} \\ \text{the best} \end{array} \right\}$  possible candidate.

Analyses for these cases can be framed around ellipsis, along the lines first proposed in Larson (2000), with structures like (49):

- (49) We interviewed the best candidate possible for us to interview.

There is a satisfying account built from standard assumptions about superlatives in Romero (2013).

It will be the other two classes that will be of interest here. These are what I'll call the weak-quantifier class, which includes *whole* and *unknown* and which permits external readings with weak quantifiers, and what I'll call the no-quantifier class, which includes *occasional* and *average* and permits external readings only with non-quantificational determiners. Of course, describing various particular determiners as 'non-quantificational' is already a bit tendentious—though for the moment, I mean this only descriptively, in the sense of Heim (1982), Kamp (1981), and DRT more generally—so more needs to be said for explicitness.

It goes beyond the scope of this paper to advocate a particular theory of how determiner quantification works in general. All we require is some general conceptual machinery to characterize particular classes. I'll refer to *every* and *most* DPs as strong and inherently quantificational; definite descriptions and other DPs that arguably directly refer as strong but not inherently quantificational; and all others as weak.

Setting the ellipsis class aside, all nonlocal readings observe a generalization:

(50) STRONG QUANTIFIER RESISTANCE GENERALIZATION

Strong, inherently quantificational determiners (*every*, *most*) are incompatible with nonlocal readings.

This has been observed for specific lexical-semantic families of adjectives, but the important point is that it seems to be true of all of them.

As we've seen, a few nonlocal adjectives—*occasional*, *average*, and *wrong*—are even more constrained in that they are incompatible with any determiner apart from (some combination of) *the*, *a*, bare plurals, and generic *your*. Stating it more officially:

(51) BROADER QUANTIFIER RESISTANCE GENERALIZATION

Some adjectives with nonlocal readings idiosyncratically resist all inherently quantificational determiners.

These generalizations are the crucial element in the taxonomy, so it may help to summarize things in a table:

(52)

	strong quant. ( <i>every, most</i> )	weak ( <i>three, many</i> )
<i>occasional</i>	✗	✗
<i>average</i>	✗	✗
<i>wrong</i>	✗	✗
<i>same</i>	✗	✗
<i>whole</i>	✗	✓
<i>unknown</i>	✗	✓
<i>inevitable</i>	✗	✓
<i>unlikely</i>	✗	✓
<i>different</i>	✗	✓
<i>possible</i>	✓	✗

Of course, the challenge now is to explain these generalizations. That's a tall order, inasmuch as it requires a synthesis of a vast array of adjectives and (collectively) a vast literature and set of analytical approaches. This won't happen in any single paper. Nevertheless, having framed the challenge in this way, we are in a better position to assess what an explanation might look like.

## 4 Some Background

### 4.1 Incorporation

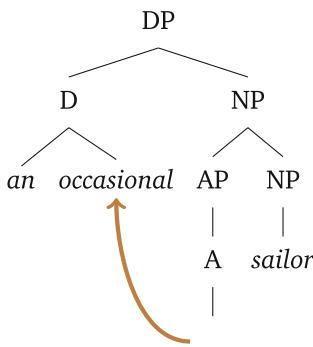
First, we must dispense with a straw man. One might imagine that external readings of adjectives are brought about simply by moving the adjective from its base position to an adverbial position, where it is interpreted as an adverb. The idea is a natural one, and I'll argue that in a certain sense it's not entirely wrong—but formulated in this crude way, it's unenlightening. Why should this movement happen? Why would an adjective have an adverb meaning? How does this help us understand the interaction of the adjective with the determiner?

More enlightening alternatives are available. There are many analyses on the market of individual instances of the larger problem of nonlocal readings, but they aren't straightforwardly generalizable to the full range of facts. There is one idea, though, that constitutes an excellent starting point. It's Larson (1999)'s proposal (further developed in Zimmermann 2000, 2003) that, in the *occasional* construction, the adjective moves from its base position to incorporate into the determiner in a process of ‘complex quantifier formation’<sup>5</sup>:

---

<sup>5</sup>I use ‘incorporation’ here following Larson and Zimmermann, in the generalized sense derived from Baker (1985) that is standard in the generative syntactic literature.

(53)



This movement creates a single quantificational determiner, *an+occasional*. It is then possible to provide this determiner with a denotation, listed in the lexicon just like that of any other. The advantage of that is that it's straightforward to capture various idiosyncrasies. If we need to stipulate that for *occasional* and *average*, the denotations of *the*, *a*, and *your* should be identical but for *wrong* they shouldn't be, we can reflect it directly. Indeed, we should *expect* such idiosyncrasies, inasmuch as the lexicon is, after all, a repository of the idiosyncratic.

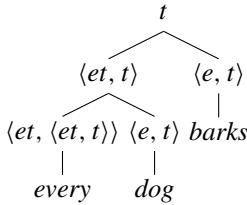
What's less comfortable is that we have to stipulate not just that *an+occasional*, *the+occasional*, and *your+occasional* all have identical denotations, but also to make precisely the same stipulation independently for *a+sporadic*, *the+sporadic*, and *the+sporadic*—and indeed for other combinations of *a*, *the*, and *your* with adjectives of this class (though see Zimmerman 2003) for some inroads on this).<sup>6</sup>

Some analysis is necessary of why these readings fail to occur with determiners other than *a*, *the*, and *your*. On this approach, it would simply be to fail to stipulate any complex determiners that fail to have these as components. It would be essentially an accidental lexical gap, a mere accident of the development of language.

This approach provides help in one way right off the bat. Quantificational determiners have access to the VP by perfectly ordinary means: Quantifier Raising (May 1977, 1985; Heim and Kratzer 1998). A generalized quantifier—the type of expression a quantified nominal denotes—takes a VP as its argument. The basic architecture of a quantified sentence is as in (54):

<sup>6</sup>This isn't uniformly a flaw. Certain combinations of frequency adjectives and determiners do seem to lack external readings for mysterious reasons. *The odd sailor strolled by* gets an external reading, but it's far more difficult to get it for *?An odd sailor strolled by*, as Gehrke and McNally (2015) observe. I'm not entirely sure what to make of these facts, but they don't strike me as sufficient reason to give up on the cause of trying to derive these generalizations from something deeper. In this specific case, the independently pragmatic naturalness of the internal reading may be relevant.

(54)



$$(55) \llbracket \text{every dog} \rrbracket = \lambda Q_{\langle e, t \rangle} . \forall x[\mathbf{dog}(x) \rightarrow Q(x)]$$

The determiner *every* here has ‘access’ to the VP in the sense that its denotation asks for a predicate,  $Q$  in (55), that it can subsequently manipulate. The manipulation of VP meanings is the signature property of adverbials, of course, so on the incorporation view, what makes it seem like *occasional* has an ‘adverbial’ external reading is that it incorporates into a quantificational determiner and therefore has access to a VP meaning. Its access to clausal material external to the DP is a side-effect of the access the VP it has by actually being, in a deeper sense, a determiner.

If an adjective is part of a quantificational determiner meaning, it will gain access to the VP as a matter of course.

Thus this approach accounts for the adverbial scope of *occasional* and its kin, for the idiosyncratic interpretations of determiners in this construction, and (by stipulation) for restrictions on the determiner. It also accounts for the restriction on coordination: any adjective in a coordinate structure would be unable to move out of it without violating the Coordinate Structure Island. In general, movement from outside of one conjunct in a coordinate structure is not possible:

- (56) a. Floyd ate rice and beans.  
 b. \*Beans<sub>1</sub>, Floyd ate rice and  $t_1$ .

That’s precisely the sort of movement that, on this view, would be required in (57) to achieve the impossible external reading:

- (57) a. The occasional and angry sailor strolled by.  
 b. #[The+occasional<sub>1</sub>] [ $t_1$  and angry] sailor strolled by.

The obligatory high position of the adjective is explained as well—any adjectives above it would block its path to the determiner.

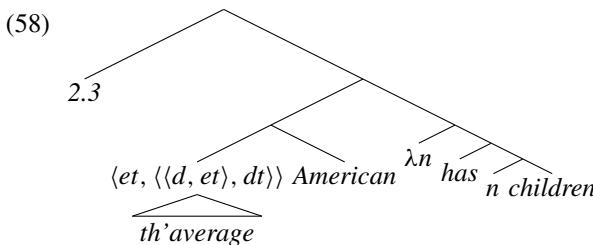
The incompatibility of external readings with degree modification would also be expected, because only a bare adjective, and not a phrasal constituent, can do head-to-head movement, the kind required here. *Occasional* on its own is the head of an AP, but *very occasional* is not. This approach may even shed light on Zimmermann (2003)’s observation that external readings are often absent where Quantifier Raising is blocked. This analysis can be extended to *average*, *wrong*, perhaps *same*, and maybe others.

Nevertheless, one might have some qualms. The movement required would seem to violate the Head Movement Constraint (Travis 1984), which would normally prevent a head from moving outside of an adjoined phrase (the AP, in this case) as in (53).

More worrying, perhaps: why are *a*, *the*, and *your* alone the determiners that have been targeted for complex quantifier formation? Could it in principle have been any other combination? And why is it that the denotations of these complex determiner-adjective combinations aren't unpredictable? If they're specified in the lexicon, one might imagine virtually arbitrary variation, but the generalizations we would like to explain aren't arbitrary. Whatever the answers to these questions, more would have to be said to make weak-determiner-compatible adjectives such as *whole*, *unspecified*, and *different* fit in.

## 4.2 Structure Versus Ontology: The First Step

Framing the current project as a trade-off between structure and ontology, at least with respect to *average*, is as I've said not novel. What I propose here is a variation on a theme from Kennedy and Stanley (2009). They observe the connection between *average* and *occasional*, and that this connection affords an analytical opportunity. For them, *average* incorporates into the determiner, just like *occasional* does for Larson (2000). The actual combinatorics required to achieve the necessary readings are complicated in ways that can be set aside, but they require a non-standard scope-taking mechanism that Barker (2007) dubbed Parasitic Scope, though appeals to it without the brand name can be found in Sauerland (1998) and earlier. The structure they propose is this:



The variable  $n$  here ranges over real numbers, or what number terms like 2.3 denote. The denotation is built up using the complex determiner *th'average* as in (59):

$$(59) \llbracket \text{th'average} \rrbracket (\llbracket \lambda n \lambda x . \text{has } n \text{ children} \rrbracket) (\llbracket \text{American} \rrbracket) (\llbracket 2.3 \rrbracket)$$

The denotation of *th'average* applies to three arguments. The first is a relation between numbers and individuals that have that number of children. The second is a property indicating what population is being averaged over, in this case, Americans. The final one is a real number indicating the computed average.

The details of implementation won't be crucial here, but they involve computing a mean on the basis of the maximal number of children each individual has,<sup>7</sup> and  $|P|$  should be interpreted as the number of individuals that have the property  $P$ :

$$(60) \quad \llbracket \text{th}'\text{average} \rrbracket = \lambda P_{(e, t)} \lambda f_{(e, nt)} \lambda n . \frac{\sum\limits_{P(x)} \mathbf{max}\{n | f(x)(n)\}}{|P|}$$

The most important point, for current purposes, is that on this view *average* DPs don't refer to anything metaphysically exotic because they don't refer to anything at all. Rather, they have an exotically high semantic type, which, coupled with incorporation from an adjective into a determiner and an unusual scope-taking operation, add up to a semantics that yields the right reading. For them, the right reading is strictly adverbial. It's the reading that can be paraphrased 'on average, Americans have 2.3 children'. It's worth noting, though, that this analysis has many of the same costs as the basic incorporation analysis, including having to stipulate the equivalence of *the*, *a*, and *your* on this reading.

### 4.3 *The Kind Analysis of 'Occasional'*

The balance between the compositional semantics and the ontology is tilted in precisely the other direction in Gehrke and McNally (2010, 2015), building on Schäfer (2007). The distinctive property of *occasional* nominals, for them, is not in their grammar but rather in their referential properties—and it is therefore there that we should locate an analysis. So they seek a simpler syntax-semantics and a richer ontology.

It would require navigating quite a bit off my intended course to do justice to their proposal, but at its heart is an idea on which I will build: kind reference. The observation is that *the occasional sailor* involves reference to realizations of sailor-kinds. Very approximately, the truth conditions of the now-familiar sailor sentence can be rendered as in (61):

- (61) The occasional sailor strolled by.

Approximately: 'Suitably-distributed realizations of strolling-by event kinds involved realizations of the sailor kind.'

The major advantage to this strategy is that it doesn't require the compositional backflips that the incorporation analysis—and especially the Kennedy and Stanley (2009) variant for *occasional*—requires. Indeed, because there is no movement at all, it doesn't violate the Head Movement Constraint. It also provides insight into why *a*, *the*, and *your* should be the determiners that uniquely have a special status in

---

<sup>7</sup>The maximality operator is required because anyone with three children also has two.

this construction. This is precisely the class of determiners that have a special status with respect to kinds and genericity:

- (62) a. The domestic dog is a good friend.
- b. A dog is a better friend than a cat.
- c. Your purple-breasted snicklewarbler is a magnificent bird. *(dialectal)*

To the extent that this approach is successful, it requires no special stipulations about the denotations of determiners. And because of that, it helps explain why determiner interpretations don't vary freely. No special stipulations are necessary to explain why *your+occasional* or the unattested *\*every+occasional* don't just happen by chance to mean something they don't actually mean.

The main shortcoming of this approach, from the current perspective, is that it's not clear how to make it scale up. On its own, it seems convincing that kind-reference is going to be a crucial ingredient in the analysis of external readings. But it's not clear to me how to make it the principal ingredient in a fully general theory.

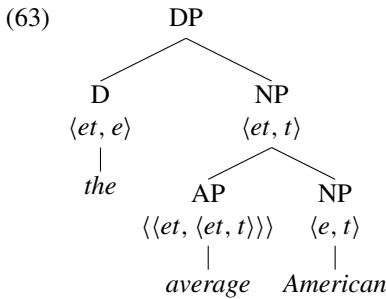
## 5 The Modular Strategy

### 5.1 Determiner-Like Adjectives

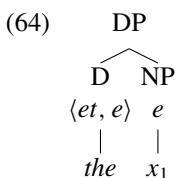
The aim of this paper is not to present a general theory of nonlocal readings, but taking a confident step in that direction requires a theory of how they arise that is modular: that is, one that relies on multiple interacting parts to arrive at an explanation. Such a theory makes it possible to activate or deactivate certain of these components to explain variation among subclasses of adjectives and—most directly at issue here—to explain the biggest split among nonlocal readings, the one between adjectives that give rise to Broader Quantifier Resistance and those that don't. (This sets aside, of course, the *possible* ellipsis class.)

One satisfying aspect of the incorporation analysis sketched above is that it reflects that nonlocal adjectives aren't prototypically adjective-like, even on a purely descriptive level. They don't pass standard diagnostics for adjectives, such as the ability to occur in comparatives, with degree modifiers, or in the complement position of *seem*. They don't conjoin with adjectives. Nor do they occur in the same positions as adjectives generally; rather, they are obligatorily high.

This might suggest incorporation or another form of syntactic differentiation, but all these properties also follow from simply assuming that nonlocal adjectives have an unusual semantic type. In the spirit of the incorporation approach, I'll assume these adjectives have precisely the same type of denotation as quantificational determiners, namely type  $\langle et, \langle et, t \rangle \rangle$ . Switching back to *average American*, the picture would be as in (63):



This has as a consequence that the node above the adjective, the NP *average American*, would denote a generalized quantifier. Following standard assumptions (see Heim and Kratzer 1998 for a review), it would therefore have to quantifier-raise and adjoin to the clause to avoid a type clash. I'll leave aside what happens higher in the clause for the moment to focus on the DP. The trace this movement leaves behind would standardly denote an individual. To make these LFs slightly easier to read later in the paper, I'll write it as a variable rather than a trace:



But this is hardly any help at all. It just gives rise to a different type clash: the NP would now denote an individual, but *the* is of type  $\langle et, e \rangle$  and expects a property.

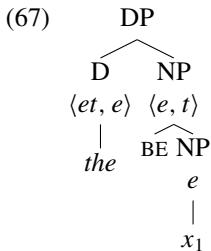
There is a natural solution. It's to adopt the standard BE type shift (Partee 1987), which shifts an individual to the property of being that individual:

$$(65) \llbracket \text{BE} \rrbracket = \lambda x \lambda y [x = y]$$

Applied to Floyd, for example, this shift would yield the property of being Floyd:

$$(66) \llbracket \text{BE} \rrbracket (\text{Floyd}) = \lambda y [\text{Floyd} = y]$$

Partee used it for copular constructions, and it has subsequently proven useful elsewhere. In this case, this resolves the type clash by providing *the* with the property-denoting argument it seeks in (64):



(68)  $\llbracket \text{BE} \rrbracket (\llbracket x_1 \rrbracket) = \lambda y[x_1 = y]$

But as it turns out, at the next node up, this shift will achieve for us something more.

## 5.2 Determiners That Work

One of the things we would like to explain is why *the*, *a*, and *your* seem to work robustly with a number of nonlocal adjectives, and why distinctions in their interpretations seem to be neutralized in the presence of frequency adjectives and *average/typical*. That result follows from the type shift alone. There is one and only one individual that has the property of being Floyd, and it is Floyd. For this reason, *the person who is Floyd* and *Floyd* mean the same thing. So too, here *the* would combine with the property the shifted trace denotes to yield the unique individual that is identical to the one the unshifted trace denotes:

- (69) a.  $\llbracket \text{the} \rrbracket = \lambda P_{\langle e, t \rangle} . \lambda y[P(y)]$   
      b.  $\llbracket \text{the} \rrbracket (\llbracket \text{BE} x_1 \rrbracket) = \lambda y[x_1 = y] = x_1$

This is precisely the same individual as the one denoted by the trace alone. The effect is as though *the* were absent entirely, as though the nonlocal adjective and its NP sister had occurred in subject position on their own.

The semantically-bleached variant of *your* that occurs in e.g. *your average American* mostly amounts to a version of *the* with a slight whiff of genericity about it, which would leave us in more or less the same place (see Gehrke and McNally 2010, 2015 for more).

As for *a*, the right result follows from a simple equivalence. To say that there's a person  $x$  such that  $x$  is wearing a hat and  $x$  is Floyd is just to say that Floyd is wear a hat. The same equivalence manifests itself in (70). The standard denotation of the indefinite article in (70a) when combined with the shifted trace denotation, as in (70b), yields an expression that asks for a predicate  $Q$  and says that some individual identical to  $x_1$  satisfies  $Q$ :

- (70) a.  $\llbracket a \rrbracket = \lambda P_{\langle e, t \rangle} \lambda Q_{\langle e, t \rangle} . \exists x[P(x) \wedge Q(x)]$   
      b.  $\llbracket a \rrbracket (\llbracket \text{BE} x_1 \rrbracket) = \lambda Q_{\langle e, t \rangle} . \exists x[x_1 = x \wedge Q(x)]$

To say that there is an individual identical to  $x_1$  of which the predicate  $Q$  holds is simply to say that  $Q$  holds of  $x_1$ :

$$(71) \exists x[x_1 = x \wedge Q(x)] \Leftrightarrow Q(x_1)$$

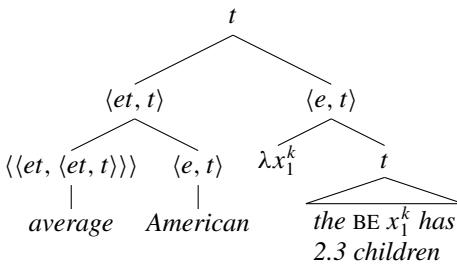
The result, again, is truth-conditionally identical to what would have happened had the determiner been absent entirely.

To articulate this a little bit further, let's adopt the toy denotation for *average* in (72a). This applies to the denotation of the modified NP, and predicates the VP meaning of the kind that corresponds to the NP meaning, using Chierchia (1998)'s  $\cap$  property-to-kind type shift<sup>8</sup>:

- $$(72) \text{a. } [\![\text{average}]\!] = \lambda P_{(e, t)} \lambda Q_{(e, t)} . Q(\cap P)$$
- $$\text{b. } [\![\text{average American}]\!] = \lambda Q_{(e, t)} . Q(\cap \text{American})$$

This probably isn't adequate on its own as a theory of *average*, and much of Kennedy and Stanley (2009) may have to be layered on top of it. A few more words on this follow in Sect. 6.1 below. But it suffices to sketch the compositional machinery. Thus the updated tree would look like this (I've ornamented the tree with a superscript  $k$  to reflect that the trace of *average American* denotes a kind):

(73)



The result of the computation is just what we need:

- $$(74) \text{a. } [\![\text{the BE } x_1^k]\!] = x_1^k$$
- $$\text{b. } [\![\text{the BE } x_1^k \text{ has } 2.3 \text{ children}]\!] = \text{has-2.3-children}(x_1^k)$$
- $$\text{c. } [\![\text{average American}]\!] = \lambda Q_{(e, t)} . Q(\cap \text{American})$$
- $$\text{d. } [\![\text{average American}]\!] ([\![\lambda x_1^k \text{ the [BE } x_1^k]\!]] \text{ has } 2.3 \text{ children}) \\ = \text{has-2.3-children}(\cap \text{American})$$

So the upshot is a semantics that requires that Americans generally have 2.3 children.

The crucial component to notice here is not the semantics of *average*, though, so much as the way the combination of the type shift, compositional assumptions, and

---

<sup>8</sup>Given this denotation, I could have equivalently dispensed with the  $\lambda Q$  in the denotation of *average* and had *average American* denote a kind directly. This is possible here only because I am using a considerably simplified denotation, though.

kind-reference have achieved the effect of ensuring that precisely the determiners that systematically license external readings yield the right result.

### 5.3 Determiners That Don't Work

What of determiners that *don't* work? Again, the nature of the movement and resulting type shift helps the situation—or rather, undermines it in the right way.

Strong determiners like *every* and *most* presuppose that their domain has more than one member. If there is only one person in the corner, for example, (75) gives rise to failure of presupposition:

- (75) Every person in the corner left.

I've spelled it out explicitly in the denotation of *every* in (76) (the colon indicates the presupposition;  $|P|$ , as before, indicates the cardinality of individuals that satisfy  $P$ )

$$(76) \llbracket \text{every} \rrbracket = \lambda P_{\langle e, t \rangle} : |P| > 1 . \lambda Q_{\langle e, t \rangle} . \forall x[P(x) \rightarrow Q(x)]$$

In (77), *every* combines with the property  $\llbracket \text{BE } x_1^k \rrbracket$ :

- (77) a. #Every average American has 2.3 children.  
       b. [average American]  $\lambda x_1^k$  [ every  $\llbracket \text{BE } x_1^k \rrbracket$  ] has 2.3 children

$$(78) \llbracket \text{BE } x_1^k \rrbracket = \lambda y[x_1^k = y]$$

But (78) is a singleton property—there is only one individual that is identical to  $x_1^k$ . It therefore violates the presupposition *every* imposes on its first argument.

This presupposition is not a peculiarity of *every*, but rather a property of strong quantificational determiners in general. Thus *most* would work similarly. Because movement below the DP level, in the framework proposed here, systematically gives rise to such singleton properties, it systematically precludes combining with strong quantifiers.

We have thus derived one of the two generalizations articulated earlier: the Strong Quantifier Resistance Generalization. All external readings observed it, so if this mechanism is crucial to deriving external readings, this explains it. Weak quantificational determiners do not have this presupposition, so they don't in general block external readings.

But what of the Broader Quantifier Resistance Generalization, the one only some adjectives observed? Some adjectives—like our test cases, *average* and *occasional*—do block the external reading with weak quantifiers too. But despite the absence of the fatal presupposition, these fail in another respect. The denotation of *three* is as in (79), a property of individuals that have a cardinality of 3:

$$(79) \llbracket \text{three} \rrbracket = \lambda x[|x| = 3]$$

When this combines with the shifted trace, it will combine intersectively with its denotation to yield (80):

$$(80) \llbracket \text{three BE } x_1^k \rrbracket = \lambda y[x_1^k = y \wedge |y| = 3]$$

This is a property satisfied by a plurality with three elements that is identical to the kind  $x_1^k$ . That means, naturally, that the kind  $x_1^k$  has to be a plurality of three elements. But kinds aren't pluralities, and they don't have cardinalities. This is pretty straightforward metaphysically, but again, linguistic evidence makes it clear. As Chierchia (1998) demonstrates especially convincingly, across languages kinds are essentially a kind of mass term. *Cheese*, for example, denotes a kind in English, and *\*three cheese* is ungrammatical.

So in this case, the problem that rules out weak quantifiers has to do with kinds, and it will be only nonlocal adjectives that leave behind kind-denoting traces that will be subject to this additional restriction. *Occasional* is also incompatible with weak quantifiers, and, as Gehrke and McNally (2010, 2015) demonstrate, its semantics also relies crucially on kinds. Nonlocal adjective with no kind overtones such as *whole* or *wrong* or *unspecified* should therefore avoid running afoul of this difficulty and be compatible with weak quantifiers even on their external readings. And indeed they are. More on both of these points follows in the subsequent two sections.

## 5.4 A Word About ‘Occasional’

*Occasional* and its kin aren't the focus here, but a brief word about how they might work in this framework is appropriate. The approach to which I'm most sympathetic would be to simply combine the insights of two competing classes of approaches. Kinds must occupy a central place, for the reasons discussed above. But quantification can play a central role too. In particular, there is no reason not to adopt the Zimmermann (2000)'s quantifier OCCASIONAL, which quantifies jointly over the individuals and events, though here it will be crucial that it be kinds and events (with  $s$  as the type of events):

$$(81) \llbracket \text{occasional} \rrbracket = \lambda P_{\langle e, t \rangle} \lambda Q_{\langle e, st \rangle}. \text{OCCASIONAL } x^k, e : {}^\cap P(x)[Q(x^k)(e)]$$

This denotation would trigger movement to a position just below where the event argument is closed, and yield sentence denotations like (82):

- (82) a. The occasional sailor strolled by.
  - b.  $\llbracket \text{occasional sailor } \lambda x_1^k \text{ the BE } x_1^k \text{ strolled by} \rrbracket$
- $$= \text{OCCASIONAL } x^k, e : {}^\cap \text{sailor}(x)[\text{strolled-by}(x^k)(e)]$$

This seems a reasonable happy medium between the two approaches.

## 5.5 The Weak Quantifier Class

There remains to discuss the class of external readings that *are* compatible with weak quantifiers. For those, though, in one sense there is little to be said. What ensured incompatibility with weak quantifiers above was the role of kinds. Adjectives whose semantics makes no special reference to kinds don't give rise to the problem of computing the cardinality of a kind.

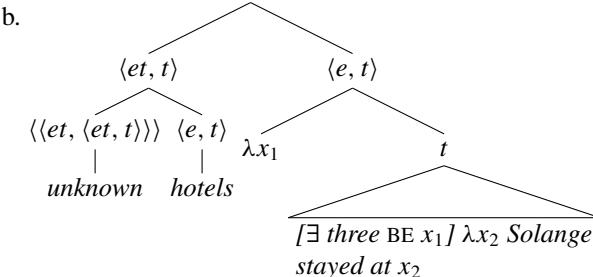
To illustrate, the denotation of *unknown* could be characterized as in (83), where I've used  $?x\phi$  to abbreviate the embedded question 'which  $x$  is such that  $\phi$ '<sup>9</sup>:

- (83) a.  $\llbracket \text{unknown} \rrbracket = \lambda P_{(e, t)} \lambda Q_{(e, t)} . \exists x [P(x) \wedge Q(x) \wedge \neg \mathbf{known} (?y[Q(y)])]$   
 b.  $\llbracket \text{unknown hotel} \rrbracket = \lambda Q_{(e, t)} . \exists x [\mathbf{hotel}(x) \wedge Q(x) \wedge \neg \mathbf{known} (?y[Q(y)])]$

What *unknown hotel* does is a little complicated. First, it requires that there exist a hotel that satisfies the property formed by raising the whole quantified NP. Second, it requires that it not be known which individuals satisfy this property.

It will help to see how this works in action. The tree for (84a) arrived at by raising would be as in (84b):

- (84) a. Solange stayed at three unknown hotels.



This assumes a null existential determiner in the head of the nominal, and that, standardly, it undergoes quantifier raising. The denotation of (84) would be as in (85):

- (85) a.  $\llbracket \text{BE } x_1 \rrbracket = \lambda y[x_1 = y]$   
 b.  $\llbracket \text{three BE } x_1 \rrbracket = \lambda y[x_1 = y \wedge |x_1| = 3]$   
 c.  $\llbracket \exists \text{ three BE } x_1 \rrbracket = \lambda g_{(e, t)} . \exists y[x_1 = y \wedge |x_1| = 3 \wedge g(y)]$   
 d.  $\llbracket \lambda x_1 [\exists \text{ three } x_1] \lambda x_2 \text{ Solange stayed at } x_2 \rrbracket = \lambda x_1 . \exists y[x_1 = y \wedge |x_1| = 3 \wedge \mathbf{stay-at}(y)(\text{Solange})] = \lambda x_1 . [|y| = 3 \wedge \mathbf{stay-at}(y)(\text{Solange})]$

<sup>9</sup>One may freely substitute one's favorite theory of indirect questions here, so far as I can see, though what I have in mind is that  $\lceil ?y[Q(y)] \rceil$  should be taken to be the set of propositions formed by varying the value of  $y$ , i.e., an abbreviation for  $\lceil \{p : \exists x[p = Q(y)]\} \rceil$ .

This is a property that holds of any three-membered plural individual such that Solange stayed at its members.<sup>10</sup>

What *unknown hotels* adds to this is that this plurality is required to consist of hotels, and that it not be known which hotels precisely these are. The computation for the full sentence is in (86):

$$(86) \quad \llbracket \text{unknown hotels } \lambda x_1 [\exists \text{ three } x_1] \lambda x_2 \text{ Solange stayed at } x_2 \rrbracket \\ = \exists x \left[ \begin{array}{l} \text{hotel}(x) \wedge |x| = 3 \wedge \text{stay-at}(x)(\text{Solange}) \wedge \\ \neg \text{known}(\exists y[\text{stay-at}(y)(\text{Solange})]) \end{array} \right]$$

The result, correctly, is that there must be three hotels at which Solange stayed, and it must not be known which hotels these are.

The crucial element in all this, though, is that there is nothing about *unknown* that prevents cardinalities from being computed, and so nothing that resists, in this instance, *three*, and more broadly any of its kin.

## 5.6 Summary

The result, then, is that there is no need for incorporation. The external scope facts follow from quantifier raising. The interpretation of determiners is standard. Restrictions on determiners follow from independent considerations. The general resistance of nonlocal adjectives to strong quantifiers follows from the compositional circumstances of their movement, which invoke a type shift with which they are incompatible. The resistance of certain nonlocal adjectives to weak quantifiers follows from independent facts about the lexical semantics of the adjective—specifically, having a kind-based semantics. Other restrictions, like the lack of coordination with ordinary adjectives and absence of degree modifiers, follow from the quantifier type of these expressions.

This means it was not necessary to stipulate which determiners support incorporation and which don't, or what interpretations result for every combination. Nor was it necessary to stipulate why *the*, *a*, and *your* wind up making the same semantic contribution, or to do so repeatedly for each frequency adjective. It also wasn't necessary to stipulate anything about the interaction of quantificational force with external readings. This is possible in part precisely because what I have offered here is only a sketch. The devil, as always, is in the details. But I hope this illustrates an analytical approach to these facts that might hope to scale up to the broader analytical picture I sought to draw.

---

<sup>10</sup>I set aside questions of distributivity and collectivity here.

## 6 Taking Stock

### 6.1 Could Things Be so Simple?

One issue remains strikingly unresolved. I've characterized the denotation for *average* I've provided above as a toy denotation. I've said, perhaps a bit defensively, that things couldn't possibly be so simple. Surely, it couldn't suffice to say that *the average American* means, essentially, the same thing as the kind-denoting nominal *Americans*, and (87a) and (87b) mean more or less the same thing:

- (87) a. The average American has 2.3 children.
- b. Americans have 2.3 children.

But the truth is, I think the simple toy version of the facts may be onto a deeper grammatical intuition than the more complicated one.

To be sure, we have the option of layering on components of the Kennedy and Stanley approach here, introducing elements of their machinery on top of the bits I propose to achieve their desired adverbial reading. There is a danger of redundancy, though. And the more one does that, the farther one gets to the connection to kind-reference, for which Gehrke and McNally provide ample evidence.

The defense of the naïve theory proceeds in several steps. The first is empirical. Suppose we adopt a theory that involves computing a mean. On such a view, (88a) and (88b) would both be predicted to be true, and, therefore, quite probably (88c) too:

- (88) a. The average human has one ovary.
- b. The average human has one testicle.
- c. The average human has one ovary and one testicle.

Yet they are all false, or in any case false outside of exceptionally odd statistical contexts. Any theory that revolves primarily around calculation of means would fail to predict this. But in a theory that relies on kind reference, it's expected. On such a view, it's the 2.3 *children* case that's puzzling.

That, I think, is precisely where we should *want* to be puzzled—that is the case that we should treat as exotic rather than as the core example. Most languages through most of human history had no reason to refer to fractions. Moreover, the semantics of fractions is independently puzzling. They are problematic completely independent of their role in *average* sentences. It makes sense, then, that the theory of *average* shouldn't be itself founded on this independent mystery.

That said, nothing in the general conception of external readings proposed here rests above all on any particular assumptions about kinds. Perhaps other notions could do the necessary work without putting us on thin ice with respect to sentences containing fractions. Indeed, I consider one such possibility in Sect. 6.2 below. The only crucial role kinds play here is to rule out computing cardinalities, which in turn

is crucial to distinguishing the weak-quantifier class from the no-quantifier class. That's not nothing, but there may be other means of accomplishing this. Nevertheless, it's worth recognizing that there are several converging lines of evidence that point to kinds or some form of genericity in these sentences: initial intuitions about what *average* sentences mean, the judgments in (88), the role kinds play in distinguishing classes of external readings, and its place in correctly predicting which determiners have which readings. One might be still able to explain *2.3 children* by simply adopting, with Kennedy and Stanley, an extraordinarily high type, but it seems right that special stipulations should be required there and not elsewhere.

It's worth pointing out, though, that one could also follow in the spirit of Carlson and Pelletier (2002) and appeal to fictive entities in place of some form of kind. This analytical avenue may actually be more available on this approach. Kennedy and Stanley argue against the fictive entity approach in part on the grounds that it doesn't explain the limited inventory of determiners possible with *average*. Those facts, however, can be explained independently here. But again, if the relevant notion of fictive entities can emerge with an appropriate kind flavor, that seems preferable on independent grounds to the alternative.

None of that directly addresses what the semantics of *2.3 children* should be. My suspicion is that an ultimately satisfying answer requires not just a theory of nonlocal readings of adjectives, but a better theory of mathematical language, and in particular of what I elsewhere call ‘semantic viruses’ (Morzycki 2017), in the spirit of Sobin (1997) syntactic viruses (see also Lasnik and Sobin 2000; Schütze 1999). I argue there that some expressions associated with educated, often highly self-conscious language may use special semantic mechanisms not otherwise available in the semantics. Making this distinction may help us distinguish which operations and what grammatical phenomena truly are exotic and may call for some brute-force high-type complexity, and where we should seek simplicity, even occasionally in the face of apparent counterevidence.

## 6.2 *Kinds and Concepts*

Sebastian Löbner (p.c.) suggests that a number of the restrictions on external readings of *average* and *occasional* may involve characterizing more precisely the concept types they give rise to. *Average American* on the relevant reading isn't a sortal concept—one that supports counting and is neither uniquely referential nor relational. That accounts for its incompatibility with strong quantifiers (#*every average American*), and perhaps for its incompatibility with stacked or conjoined adjectives (#*an average (and irritable American)*).

This mode of explanation in some respects has the same shape as a kind-based one, or indeed as one organized around fictive entities. They all seek to derive the properties of the expression from the ontological status of the extension of the nominal. Both kinds and the relevant non-sortal concepts are uncountable. It doesn't seem too far-fetched to claim that fictive entities might not be countable either, though

that's less obvious. Insomniacs are sometimes advised to count sheep in order to fall asleep, yet under normal circumstances the livestock in one's bedroom are entirely fictive. Likewise, the resistance to quantification that I earlier attributed to a failure of presupposition could be attributed to countability as well. As I expressed it in (76), the presupposition involved determining the cardinality of individuals that satisfy the property expressed by the nominal argument. In this implementation, that is not undefined. Even though this property has in its domain kinds, it denotes a property that holds of precisely one kind. Therefore it is countable. This follows from how the movement and type shift interact. One might imagine, though, an alternative analysis where the inherent countability of the noun is crucial. In order for the analysis of adjective stacking and conjunction to go through, however, one really would have to have the NP *average American* denote this concept kind quite low in the tree, before any type shifts have taken place. On this view, then, the crucial difference could be viewed as being in how high in the structure of the nominal kinds are invoked. But there are good reasons to think properties of kinds are to be found deep in the nominal extended projection, very near the noun (Zamparelli 1995 among others). So this fact too may be insufficient to distinguish these two approaches on a deep level, setting aside particular analytical choices I've made here.

The adjective order facts, however, might be of use. Most evidence for a layer in the nominal projection that is concerned with kinds rather than objects suggests that it is the lower of the two. So-called Bolinger contrasts (Bolinger 1967; see Morzycki 2016a and Leffel 2014 for extensive discussion) such as the one in (89) show that adjectives lower in the nominal ascribe inherent or individual-level properties, and higher ones ascribe contingent or stage-level properties:

- (89) a. the invisible visible stars  
      b. #the visible invisible stars

On its only possible reading, (89a) refers to stars that are visible in principle but invisible at the moment, perhaps by clouds. But (89b) is contradictory, because it refers to stars that are invisible in principle but visible at the moment. A broadly similar fact, in the spirit of Larson (1998, 2000):

- (90) an ugly beautiful dancer  
      a. 'an ugly person who dances beautifully'  
      b. \*'a beautiful person who dances in an ugly way'  
  
  (91) a beautiful ugly dancer  
      a. \*'an ugly person who dances beautifully'  
      b. 'a beautiful person who dances in an ugly way'

Larson marshals such facts to argue for a generic quantifier in the nominal projection. But be it about kinds or not, the domain of genericity in the nominal is low. Yet as we've seen, adjectives associated with external readings are exclusively high. A reminder:

- (92) An ugly occasional sailor strolled by.
- a. ‘An ugly person who sails occasionally strolled by.’ *(internal)*
  - b. \*‘Occasionally, an ugly sailor strolled by.’ *(external)*
- (93) An occasional ugly sailor strolled by.
- a. \*‘An ugly person who sails occasionally strolled by.’ *(internal)*
  - b. ‘Occasionally, an ugly sailor strolled by.’ *(external)*

If kinds or non-sortal properties were at issue lower in the nominal, this effect would be expected to be either reversed or absent entirely.

One appeal of such an approach, in either of these incarnations, would be that the quantificational facts and the facts about conjunction and stacking could be brought under the same rubric. As it stands, the latter derive from the quantificational type of the NP. A major disadvantage is that they wouldn’t readily extend to the rather large class of adjectives compatible with weak quantifiers. Nor, in the absence of a scope-taking mechanism, would they permit the adjective access to the VP denotation. Yet this access is precisely what seems to be required for e.g. epistemic adjectives such as *unknown*, as shown in Sect. 5.5.

## 7 Final Remarks

To close, a few words about the commonly expressed intuition that nonlocal readings are a grammatical oddity. These adjectives are indeed odd, but in a precise and interesting sense. They are odd in the way that platypuses and lungfish are odd: they are—perhaps metaphorically, or perhaps more than metaphorically—transitional forms in an evolutionary progression, unusual because they combine features of two distinct categories that we normally regard as mutually exclusive. Over succeeding generations of speakers, certain adjectives may emerge from the swampy depths of the inner NP to which they are usually confined, and tentatively make their way onto the dry land of the determiner domain. They can’t be expected to make this leap in a single stride, so we can observe them in the midst of their evolutionary journey and thereby discover more about both their evolutionary origin and their destination. Like platypuses and lungfish, they are important and analytically revealing not despite their strangeness, but because of it.

Substantively, the proposal was that nonlocal adjectives have quantificational determiner denotations, trigger raising of the NP in which they occur, stranding the determiner, and sometimes require properties of kinds as their arguments. This isn’t a general theory of all nonlocal readings, naturally. That would be far too ambitious for any single paper. But it has the shape of a general theory, and my hope is that further research will be able to fill in the gaps in a similar spirit.

From the broader cognitive perspective, though, one of the larger lessons is the balance between the explanatory burden on the ontology and on the structural machinery. For *average*, for example, one might have gone in the direction of recognizing

‘average Americans’ as actual, if very abstract, objects in the model, ‘fictive persons’. For *occasional*, I followed Gehrke and McNally (2010, 2015) in placing a great deal of explanatory weight on the notion of kinds, if perhaps not quite so much weight as they have.

On the other hand, structural components played a crucial role. For *average*, one could go so far as Kennedy and Stanley (2009) do, and invoke quite high-powered syntactic and semantic machinery to twist the tree into the shape we require. For *occasional*, Larson (1999), Zimmermann (2000) and others provide a path that also requires quite a bit of syntactic machinery.

It is misguided, I think, to ask where we wind up in each respect: how much compositional structure do we need, how much metaphysics, and what the right balance is. Rather, we should recognize that there may be some explanatory trade-offs, but that inevitably, we will need a bit of both modes of explanation—and it is up to language to tell us how much we need of either.

**Acknowledgements** I’ve presented various portions and versions of this work to various people and audiences. Thanks to Adam Gobeski, Ai Kubota, Ai Taniguchi, Anna Szabolcsi, Anne-Michelle Tessier, Barbara Partee, Berit Gehrke, Bernhard Schwarz, Cara Feldscher, Chris Barker, Curt Anderson, Daniel Gutzmann, Friederike Moltmann, Gabriel Roisenberg Rodrigues, Galit Sassoon, Haley Farkas, Hannah Forsythe, Henry Davis, Hotze Rullmann, Irene Heim, Josh Herrin, Kay Ann Schlang, Kyle Rawlins, Lisa Matthewson, Lucas Champolion, Manfred Krifka, Norbert Hornstein, Omer Preminger, Paul Pietroski, Rose-Marie Déchaine, Sebastian Löbner, Stephanie Solt, Taehoon Hendrik Kim, Yan Cong, and audiences at the University of Maryland, Sinn und Bedeutung in Tübingen, Boston University, New York University, the University of British Columbia, and of course the Cognitive Structures conference. Berit and Sebastian kindly provided written comments on an earlier draft of this paper, and I’m especially grateful to Sebastian for involving me in the intellectual community around Düsseldorf.

## References

- Abusch, D., & Rooth, M. (1997). Epistemic NP modifiers. In A. Lawson (Ed.), *Proceedings of Semantics and Linguistic Theory (SALT)* (Vol. 7). Ithaca, NY: CLC Publications.
- Alrenga, P. (2006). Scalar (non-)identity and similarity. In D. Baumer, D. Montero, & M. Scanlon (Eds.), *Proceedings of the West Coast Conference on Formal Linguistics (WCCFL)* (Vol. 25). Somerville, MA: Cascadilla Press.
- Alrenga, P. (2007a). Comparisons of similarity and difference. In *Proceedings of the 2005 Workshop on the Formal Analysis of Adjectives (JET Adjectifs)*. 7 Jussieu: University of Paris.
- Alrenga, P. (2007b). Tokens, types, and identity. In *Proceedings of the North East Linguistic Society (NELS)* (Vol. 38). Amherst: GLSA Publications.
- Bach, E. (1986). The algebra of events. *Linguistics and Philosophy*, 9(1), 5–16.
- Bach, E. (1989). *Informal lectures on formal semantics*. Stony Brook: State University of New York Press.
- Baker, M. C. (1985). *Incorporation: A theory of grammatical function changing*. Doctoral dissertation, MIT.
- Barker, C. (2007). Parasitic scope. *Linguistics and Philosophy*, 30(4), 407–444.
- Beck, S. (2000). The semantics of ‘different’: Comparison operator and relational adjective. *Linguistics and Philosophy*, 23(2), 101–139.
- Bolinger, D. L. (1967). Adjectives in English: Attribution and predication. *Lingua*, 18, 1–34.

- Brasoveanu, A. (2011). Sentence-internal different as quantifier-internal anaphora. *Linguistics and Philosophy*, 34, 93–168.
- Büring, D. (2007). More or less. In *Proceedings of the Chicago Linguistics Society (CLS)* (Vol. 43).
- Carlson, G., & Pelletier, F. J. (2002). The average American has 2.3 children. *Journal of Semantics*, 19(1), 73–104.
- Carlson, G. N. (1987). ‘Same’ and different: Some consequences for syntax and semantics. *Linguistics and Philosophy*, 10(4), 531–565.
- Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics*, 6(4), 339–405.
- Cinque, G. (2010). *The syntax of adjectives: a comparative study. Linguistic inquiry monographs* (Vol. 57). Cambridge, MA: MIT Press.
- DeVries, K. (2010). *Frequency adjectives*. BA thesis, Michigan State University.
- Gehrke, B., & McNally, L. (2010). Frequency adjectives and assertions about event types. In E. Cormany, S. Ito, & D. Lutz (Eds.), *Proceedings of Semantics and Linguistic Theory (SALT)* (Vol. 19, pp. 180–197). Ithaca, NY: eLanguage.
- Gehrke, B., & McNally, L. (2015). Distributional modification: The case of frequency adjectives. *Language*, 91(4), 837–870.
- Haïk, I. (1985). *The syntax of operators*. Doctoral dissertation, MIT.
- Hall, J., & Robert, A. (1973). The transferred epithet in P. G. Wodehouse. *Linguistic Inquiry*, 4(1), 92–94.
- Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. Doctoral dissertation, University of Massachusetts, Amherst.
- Heim, I. (1985). *Notes on comparatives and related matters*. Ms., University of Texas, Austin.
- Heim, I. (2006). Little. In M. Gibson & J. Howell (Eds.), *Proceedings of Semantics and Linguistic Theory (SALT)* (Vol. 16). Ithaca, NY: CLC Publications.
- Heim, I. (2008). Decomposing antonyms? In A. Grønn (Ed.), *Proceedings of Sinn und Bedeutung* (Vol. 12, pp. 212–225). Oslo: ILOS.
- Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Oxford: Blackwell Publishing.
- Hornstein, N. (1984). *Logic as grammar*. Cambridge, MA: MIT Press.
- Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, & M. Stokhof (Eds.), *Formal methods in the study of language*, Part 1 (Vol. 135, pp. 277–322). Amsterdam: Mathematical Centre Tracts. Reprinted In J. Groenendijk, T. Janssen, & M. Stokhof (Eds.). (1984). *Truth, interpretation, and information; Selected Papers from the Third Amsterdam Colloquium* (pp. 1–41). Dordrecht: Foris.
- Keenan, E. L. (1992). Beyond the frege boundary. *Linguistics and Philosophy*, 15(2), 199–221.
- Kennedy, C. (1997). *Projecting the adjective: The syntax and semantics of gradability and comparison*. Doctoral dissertation, UC Santa Cruz. Published in 1999 by Garland, New York.
- Kennedy, C. (2001). Polar opposition and the ontology of degrees. *Linguistics and Philosophy*, 24(1), 33–70.
- Kennedy, C., & Stanley, J. (2009). On ‘average’. *Mind*, 118(471), 583–646.
- Larson, R. (1998). Events and modification in nominals. In D. Strolovitch & A. Lawson (Eds.), *Proceedings of Semantics and Linguistic Theory (SALT)* (Vol. 12). Ithaca, NY: CLC Publications.
- Larson, R. (1999). *Semantics of adjectival modification. Lecture notes*. Amsterdam: LOT Winter School.
- Larson, R. K. (2000). ‘ACD’ in AP? In R. Billerey & B. D. Lillehaugen (Eds.), *Proceedings of the West Coast Conference on Formal Linguistics (WCCFL)* (Vol. 19). Somerville, MA: Cascadilla Press.
- Lasersohn, P. (2000). Same, models and representation. In B. Jackson & T. Matthews (Eds.), *Proceedings of Proceedings of Semantics and Linguistic Theory (SALT)* (Vol. 10, pp. 83–97). Ithaca, NY: CLC Publications.
- Lasnik, H., & Slobin, N. (2000). The who/whom puzzle: On the preservation of an archaic feature. *Natural Language and Linguistic Theory*, 18(2), 343–371.

- Leffel, T. (2014). *The semantics of modification: Adjectives, nouns, and order*. Doctoral dissertation, New York University.
- Löbner, S. (2000). Polarity in natural language: Predication, quantification and negation in particular and characterizing sentences. *Linguistics and Philosophy*, 23(3), 213–308.
- Majewski, H. (2002). *The same and different: The same or different?*. Ms., University of Massachusetts, Amherst.
- May, R. (1977). *The grammar of quantification*. Doctoral dissertation, MIT, Cambridge, MA.
- May, R. (1985). *Logical form*. MIT Press.
- Moltmann, F. (1992). ‘Reciprocals and same/different’: Towards a semantic analysis. *Linguistics and Philosophy*, 15(4), 411–462.
- Moltmann, F. (1997). Intensional verbs and quantifiers. *Natural Language Semantics*, 5(1), 1–52.
- Moltmann, F. (2005). Part structures in situations: The semantics of individual and whole. *Linguistics and Philosophy*, 28(5), 599–641.
- Morzycki, M. (2002). Wholes and their covers. In B. Jackson (Ed.), *Proceedings of Semantics and Linguistic Theory (SALT)* (Vol. 12). Ithaca, NY: CLC Publications.
- Morzycki, M. (2016a). *Modification. Key topics in semantics and pragmatics*. Cambridge: Cambridge University Press.
- Morzycki, M. (2016b). Toward a general theory of nonlocal readings of adjectives. In N. Bade, P. Berezovskaya, & A. Schöller (Eds.), *Proceedings of Sinn und Bedeutung* (Vol. 20).
- Morzycki, M. (2017). Some viruses in the semantics. In N. LaCara, K. Moulton, & A.-M. Tessier (Eds.), *A Schrift to Fest Kyle Johnson*. University of Massachusetts Linguistics Open Access Publications.
- Nunberg, G. (1984). Individuation in context. In M. Cobler et al. (Eds.), *Proceedings of the West Coast Conference on Formal Linguistics (WCCFL)* (Vol. 2, pp. 203–217). Stanford: Stanford Linguistics Association.
- Partee, B. (1987). Noun phrase interpretation and type-shifting principles. In J. Groenendijk, D. de Jongh, & M. Stokhof (Eds.), *Studies in discourse representation theory and the theory of generalized quantifiers* (pp. 115–143). Dordrecht: Foris (reprinted in Partee Parter (2002)).
- Partee, B. H., & Portner, P. (Eds.). (2002). *Formal semantics: The essential readings*. Oxford: Blackwell Publishing.
- Romero, M. (2013). Modal superlatives: A compositional analysis. *Natural Language Semantics*, 21, 79–110.
- Sauerland, U. (1998). *The meaning of chains*. Doctoral dissertation, MIT, Cambridge, MA.
- Schäfer, R. (2007). On frequency adjectives. In E. Puig-Waldmüller (Ed.), *Proceedings of Sinn und Bedeutung* (Vol. 11). Barcelona: Universitat Pompeu Fabra.
- Schmitt, C. (2000). Some consequences of the complement analysis for relative clauses, demonstratives and wrong adjectives. In A. Alexiadou, P. Law, A. Meinunger, & C. Wilder (Eds.), *The syntax of relative clauses* (pp. 309–348). Amsterdam: John Benjamins Publishing.
- Schütze, C. (1999). English expletive constructions are not infected. *Linguistic Inquiry*, 30(3), 467–484.
- Schwarz, B. (2005). Modal superlatives. In E. Georgala & J. Howell (Eds.), *Proceedings of Semantics and Linguistic Theory (SALT)* (Vol. 15). Ithaca, NY: CLC Publications.
- Schwarz, B. (2006). Attributive wrong. In D. Montero, D. Baumer, & Michael Scanlon (Eds.), *Proceedings of the West Coast Conference on Formal Linguistics (WCCFL)* (Vol. 25, pp. 362–370). Somerville, MA: Cascadilla Press.
- Schwarz, B. (To appear). The wrong number: Non-local adjectival modification. In L. Matthewson, C. Meier, H. Rullmann, & T. E. Zimmermann (Eds.), *The Blackwell companion to semantics*. Blackwell Publishers.
- Sobin, N. (1997). Agreement, default rules, and grammatical viruses. *Linguistic Inquiry*, 28(2), 318–343.
- Stump, G. T. (1981). The interpretation of frequency adjectives. *Linguistics and Philosophy*, 4(2), 221–257.
- Travis, L. (1984). *Parameters and effects of word order variation*. Doctoral dissertation, MIT.

- Zamparelli, R. (1995). *Layers in the determiner phrase*. Doctoral dissertation, University of Rochester (as revised in 2000).
- Zimmermann, M. (2000). Pluractional quantifiers: The *Occasional* construction in English and German. In B. Jackson & T. Matthews (Eds.), *Proceedings of Semantics and Linguistic Theory (SALT)* (Vol. 10). Ithaca, NY: CLC Publications.
- Zimmermann, M. (2003). Pluractionality and complex quantifier formation. *Natural Language Semantics*, 11(3), 249–287.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# **Concept Theory**

# How Can Semantics Avoid the Troubles with the Analytic/Synthetic Distinction?



Roberto G. de Almeida and Caitlyn Antal

One failure a week is just bracing and good for you.  
—Allen Newell (1991).

**Abstract** At least since Quine (From a logical point of view. Harvard University Press, Cambridge, MA, 1953) it has been suspected that a semantic theory that rests on defining features, or on what are taken to be “analytic” properties bearing on the content of lexical items, rests on a fault line. Simply put, there is no criterion for determining which features or properties are to be analytic and which ones are to be synthetic or contingent on experience. Deep down, our concern is what cognitive science and its several competing semantic theories have to offer in terms of solution. We analyze a few cases, which run into trouble by appealing to analyticity, and propose our own solution to this problem: a version of atomism cum inferences, which we think it is the only way out of the dead-end of analyticity. We start off by discussing several guiding assumptions regarding cognitive architecture and on what we take to be methodological imperatives for doing semantics within cognitive science—that is a semantics that is concerned with accounting for mental states. We then discuss theoretical perspectives on lexical causatives and the so-called “coercion” phenomenon or, in our preferred terminology, indeterminacy. And we advance, even if briefly, a proposal for the representation and processing of conceptual content that does away with the analytic/synthetic distinction. We argue that the only account of mental content that does away with the analytic/synthetic distinction is atomism. The version of atomism that we sketch accounts for the purported effects of analyticity with a system of inferences that are in essence synthetic and, thus, not content constitutive.

---

R. G. de Almeida (✉) · C. Antal  
Department of Psychology, Concordia University, Montreal, QC, Canada  
e-mail: [roberto.dealmeida@concordia.ca](mailto:roberto.dealmeida@concordia.ca)

C. Antal  
e-mail: [caitlyn.antal@mail.concordia.ca](mailto:caitlyn.antal@mail.concordia.ca)

**Keywords** Analytic/synthetic distinction · Compositionality · Causatives · Indeterminacy · Coercion · Intentional fallacy · Concepts · Atomism · Cognitive architecture

Our concern in this paper is, on the surface, not new. For long—at least since Quine (1953) in modern times, to say little of Kant’s “cleavage” problems way back then—it has been suspected that a semantic theory that rests on defining features, or on what are taken to be “analytic” properties bearing on the content of lexical items, rests on a fault line. Simply put, there is no criterion for determining which features or properties are to be analytic and which ones are to be synthetic or contingent on experience. But that is just the glossy if old shell of our concern. Deep down, our concern is what cognitive science and its several competing semantic theories have to offer in terms of solution, if any at all. With this in mind, we analyze a few cases, which run into trouble by appealing to analyticity, and propose our own solution to this problem: a version of atomism cum inferences. We are aware that the proposal we have to offer is at odds with widely held views, but we think it is the only way out of the dead-end of analyticity, if one is not to be burdened with producing an analytic/synthetic criterion. We start off by discussing several guiding assumptions regarding cognitive architecture and on what we take to be methodological imperatives for doing semantics within cognitive science—that is a semantics that is concerned with accounting for mental states. We then discuss theoretical perspectives on a range of seemingly disconnected phenomena—in particular lexical causatives and the so-called “coercion” phenomenon or, in our preferred terminology, indeterminacy. And we advance, even if briefly, a proposal for the representation and processing of conceptual content that does away with the analytic/synthetic distinction. We will argue that the only account of mental content that does away with the analytic/synthetic distinction is atomism. The version of atomism we will sketch accounts for the purported effects of analyticity with a system of inferences that are in essence synthetic and, thus, not content constitutive.

## 1 Semantics and the Architecture of Cognition

It is not uncommon for cognitive scientists working in semantics to mix their metaphors regarding how they envision the nature of mental representations and processes. Perhaps they do so inadvertently, but the price is a lack of clarity on what one takes to be the very nature of the representation of *content* and the computational processes that are content-bearing. And if there is one issue that research in semantics needs to be clear about, it is how it conceives content representation and processing. As an example, consider sentence (1).

- (1) Mary began a book.

Imagine now that the issue at hand is how a sentence such as (1) might be interpreted. The proposal quoted in (2) is apropos the sorts of psychological events carried out during the comprehension process of (1). The semantic issues underlying this proposal will be dealt with a little later, but we start off with the commitments of this proposal vis-à-vis cognitive architecture.

- (2) “(a) When encountering the noun book, comprehenders access the word’s lexical entry and attempt to integrate various stored senses of this word into the evolving semantic representation of the sentence.
- (b) The mismatch between the verb’s selectional restrictions and the stored senses of the noun triggers a coercion process.
- (c) Comprehenders use salient properties associated with the complement noun and other relevant discourse elements (including but not necessarily limited to the agent phrase) to infer a plausible action that could be performed on the noun.
- (d) Comprehenders incorporate the event sense into their semantic interpretation of the VP by reconfiguring the semantic representation of the complement, converting  $[\beta \text{ began}[\alpha \text{ the book}]]$  into  $[\beta \text{ began}[\alpha \text{ reading the book}]]$ . (Conceivably, this could also require reconfiguration of an associated syntactic representation.)” (Traxler et al., 2005, p. 4)

We use this as a convenient example of the kinds of constraints—or lack thereof—that may drive semantic proposals within the language processing literature. As we will see, similar proposals abound in semantic theory.

To begin, our commitments unequivocally reside with the view that representations are symbolic, with processes over these representations being computational. These general commitments come with numerous caveats. First, it is not clear whether the nature of computations performed over symbolic representations involve hard-wired algorithmic, intra-modular kinds of principles, or heuristic, perhaps malleable principles. This difference is important for semantics because, by hypothesis, it marks the boundary between linguistically-driven computations bearing on “shallow” meaning (viz., a logical form), and those deemed pragmatic or based on world-knowledge, contingent on experience. We mentioned “intra-modular” computations because our proposal relies on there being a modular level of linguistic computations whose output is a form of compositional semantic representation, a shallow one nonetheless (see Fodor, 2001; and de Almeida, 2018; and de Almeida & Lepore, 2018, for recent discussion).

Postulating that linguistic processes are computations over symbolic representations is crucial to our take on what sorts of knowledge representation enter into tasks such as understanding a sentence or having a thought. This is so because we assume that some of these processes are executed in virtue of the formal properties of the expressions that are computed, including properties of its constituent symbols, while others are entirely dependent on the content of token symbols—or the content that token symbols point to. Furthermore, we assume that semantic units—or

*concepts*—are the very elements of higher-level representations and processes, not only of linguistic representations proper. That is, thoughts have concepts as their most elementary parts, and those happen to be the same elements one recovers in the process of understanding a sentence; they are the same we ought to use in semantic analysis. As such, we assume that in order to account for the nature of these cognitive processes—that is, in order to account for the nature of those thoughts—it is crucial we not only understand the nature of the elementary parts, but also how they combine to yield the meaning that the thought carries.

Moreover, we think that to entertain a thought is to entertain something like a *proposition* whose basic elements are concepts. We take a proposition to be a mental object, a symbolic expression standing for the meaning of a sentence or other higher cognitive representation. Thus, we argue that *any* complex representation carrying content is propositional, baring cases in which ideas are incomplete (viz., arguments are not saturated) or when representations refer to individuals.<sup>1</sup> Thinking, thus, entails combining all the elementary concepts into series of propositions, which are most likely represented as something akin to a logical form specifying the relations between conceptual constituents (see Kintsch, 1974; and McKoon & Ratcliff, 1992, for early propositional theories). This view also applies to the process of language comprehension: understanding a sentence requires recovering the meanings of words/morphemes in the context of the *proposition* that the sentence expresses. Propositions are thus the mental objects whose referents are states and events in the world (and ideas about events and states in the imaginary world, if you will). In order for propositions to refer, or in order for propositions to stand for the events and states whose contents they represent, they have to compose, and in order for them to compose they require a syntax.

Much of what we talk about in the present chapter, thus, has a particular notion of compositionality lurking in the background: namely, one that takes lexical and functional constituents and how they are combined syntactically to determine sentence-level meaning. Clearly, any position one takes on the analytic/synthetic distinction (or lack thereof) has direct consequences for the kinds of elements that enter into the composition of meaning. For instance, let us assume that one holds an enriched form of compositionality, as proposed by Pustejovsky (1995) and Jackendoff (2002)—a proposal to which (2) above adheres. Leaving details aside, enriched compositionality takes the meaning of a sentence to rely on the interpolation of some features or ontologically primitive properties stored within lexical entries. Such a view is burdened with establishing an analytic/synthetic distinction. In principle, by appealing to the internal analyses of lexical items, compositionality cannot hold, for analyticity is necessarily unbounded, thus holistic. Furthermore, assuming that our thoughts are productive, and that productivity requires compositionality, then thoughts ought to be compositional. Thus any theory on the basic elements of meaning necessarily needs to account for the compositionality of thoughts (see Fodor, 1998, for a similar

---

<sup>1</sup>We could argue that general or singular terms carry a property, viz., that ‘ $\exists x (\text{MARY} = x)$ ’ is about being Mary. But we will eschew this issue and assume that complex representations include at a minimum singular terms and their predicates.

point). We think, in summary, that holding on to a strict notion of compositionality is imperative for determining which concepts theory prevails. However, as we will see in Sect. 3, there are different approaches to compositionality and this issue interacts with the position one takes with regards to the analytic/synthetic distinction.

So far, this general view of the nature of complex representations strikes us as standard, though by no means consensus. But before we move on to discuss analyticity in semantics, we have two other brief methodological observations to make regarding semantics research in cognitive science. The first methodological observation is this: since we are realists and naturalists about mental representations—semantic or otherwise—we contend that to *do* semantics one needs to appeal to all tools of cognitive science, bar none. We take it that linguistic methods may take precedence over others, for crosslinguistic generalizations and distributional properties of expressions often provide us with rich data, supporting arguments for the reality of particular types of semantic algorithms. But by the same token, we take the experimental tools employed in cognitive psychology and neuroscience to be crucial to advance theory, rather than simply supporting linguistic postulates. As Fodor, Fodor, and Garrett (1975) once suggested, native speakers' intuitions are psychological data; and if we are tasked to investigate the realm of psychological data, experimental evidence might be at par with crosslinguistic and distributional evidence. This is important to mention here because what we are about to discuss requires analyzing certain phenomena not only in light of theoretical arguments, but also relying on the results of empirical observations typically obtained in experiments.

The second methodological observation we want to make regards how semantics research often proceeds. We take it that the fault line of the analytic/synthetic distinction, which we will address in the next section, has caused some other cracks in the foundations of semantics. Virtually all attempts to develop a theory of features has taken place by appealing to what one knows to be true about referents—objects and events—in the world, which are not necessarily the kinds of information one represents in mind about these objects and events. Appeals to intuitions here can only go so far. We surmise, however, that much of what drives the proposal for feature sets as constituents of concepts relies on what has been called the “intentional fallacy”. In a nutshell, the intentional fallacy arises when the particular properties that one assumes to be part of a stimulus are attributed to its mental representation. In psychology, this is sometimes referred to as the “stimulus error”, after Titchener (1909). The intentional fallacy permeates work in semantics, for any semantics that appeal to features has the burden of establishing the criteria for what is to be taken as true properties of a stimulus (whatever those may be) from properties that may result from one's knowledge or beliefs about that particular stimulus. To put it simply, what the researcher knows to be true about a referent is not necessarily true of its mental representation. The consequences of this fallacy are pervasive, crucially affecting the discussion on what is analytic and synthetic, and by extension, where the line should be drawn between semantics and pragmatics (for further discussion, see de Almeida, 2018). As we will see, a key issue—in line with what we see in proposal (2)—is the idea of “coercion”. We turn to these matters now.

## 2 The Analytic/Synthetic Distinction and Semantic Theories

We start off by briefly revisiting the problem of analyticity and why it poses a challenge for semantic theories—at least semantic theories that share our architectural commitments—in particular the key issue of compositionality. We do so aware that these issues are far from new. But at the same time, we are concerned that they are rarely, if ever, addressed in the semantics literature.<sup>2</sup>

The analytic/synthetic distinction has been like a dark cloud over semantics ever since Quine wrote his *Two dogmas* paper. Quine was interested in debunking a kind of semantics—in particular Carnap’s—that appealed to what Carnap called logically true (or *L-true*) as opposed to “indeterminate” or factual (*F-true*) statements. The distinction goes back at least to Kant’s attempt at separation between analytic (*L-true*) and synthetic (*F-true*) (see Carnap, 1956, Chap. 1). But as Quine showed, there were no firm criteria for establishing this difference: in essence, *L-true* and *F-true* were sourced from the same data, even if on the surface some statements appear to be true in virtue of the meaning of their constituents (the likes of *A dog is an animal*). It should be clear, before we advance discussion, that our concern is not with truly analytic statements such as those in which a conjunction entails its parts. These are run over form—something like  $P \& Q \rightarrow P$ . The first case is obviously compatible with the architecture we adopt: in fact it is essential to algorithmic cognitive processes that they run over form, not content, such that it is always the case that  $P \& Q \rightarrow P$  or  $P \& Q \rightarrow Q$ , no matter what  $P$  and  $Q$  stand for. Thus, analyticity of form holds. Our concern is with other, often subtler, forms of analyticity, common to lexical-semantic theories as well as theories of composition relying on certain types of semantic operations such as “coercion”. And, more broadly, our main concern is with the shaky ground upon which all of semantics that appeal to analytic *features* stands.

There are, we think, roughly three ways to conceive how a concept might enter into—i.e., contributes content to—a proposition. (i) The first is by contributing its full content, whatever that may be. If one believes concepts to be composed of particular sets of features, then the content that a given concept contributes to a proposition must necessarily be that particular set of features—nothing more, nothing less. (ii) Another way in which a concept might contribute content to a proposition is by contributing some, but not necessarily all, of its features. If one believes a concept to be made up by a set of features, then, the kinds of features that a concept might contribute to a particular proposition is relative to the particular context of the proposition—that is, it is sensitive to other constituent concepts, perhaps to the wider discourse, and perhaps to the syntax of the expression. And (iii) the third way in which a concept

---

<sup>2</sup>An anonymous reviewer was right at pointing out, among other problems, that the analytic/synthetic issue that we are trying to “reawaken” is “not new”. This, of course, is not an argument against our view. If anything, this is an embarrassment for semantic theories. We believe that the two case studies we discuss below, though limited in scope, are representative of a widespread practice in semantics. It should be noted that the kind of a/s issue we are raising is about mental representation, not linguistic analysis.

can contribute content to a proposition is somewhat similar to (i), but does away with analyticity: concepts contribute all their content, except that, according to this view, a concept has no features. In the present section, we will discuss (i) and (ii); the case for (iii) will be further advanced in Sect. 3.

We cannot possibly be exegetic in our evaluation of semantic theories that are committed to analyticity (see, e.g., Engelberg, 2011a, for review). Our goals here are to illustrate the state of the art and thus motivate our proposal for moving away from analyticity—namely, to make the case for our brand of atomism. And we will substantiate our case by discussing work from two particular semantic phenomena, one involving the representation of causative verbs, and one involving the representation of what we call “indeterminate” sentences, which in some circles is known as “coercion”. These two cases are illustrative for two reasons. The first, and perhaps most important one, is because both cases expose the root of the problem we want to shed light on: the problem of analyticity in semantics. The nature of the representation of causative verbs has long been the focus of disputes in linguistics and lexical semantic theories at least since the time of generative semantics (e.g., McCawley, 1972). The case of indeterminate sentences such as (1) has also received some attention early on (see Culicover, 1970). As we will see, these two topics are representative of how intuitions about meaning can lead to the intentional fallacy trap. And both represent challenges to the classical way of conceiving compositionality. But as we will see, in Sect. 3, we offer a parsimonious treatment of these two cases with the type of atomism *cum* inferences we propose and the classical notion of compositionality it entails. The second reason we focus on these two cases is, not coincidentally, that they have been topics of our own research—so we conveniently stay close to familiar cases to make a point we deem fundamental for investigating semantics in cognitive science, more broadly.

## 2.1 Causatives

Most theories of lexical semantic representation are committed to a form of analyticity that takes lexical meaning to be represented in terms of a cluster of features, usually expressed in the form of templates filled with variables and predicates. Causative verbs are the paradigm example as they have been the topic of many disputes between camps. A typical case is (3a), whose meaning is represented in (3b).

- (3) a. John<sub>x</sub> broke the vase
- b. [[x ACT] CAUSE [BECOME [y <BROKEN> ]]]

A representation such as in (3b), in the notation of lexical semantics (Levin & Rappaport Hovav, 2005) is nonetheless representative of other approaches such as conceptual semantics (Jackendoff, 1990, 2002), cognitive semantics (Croft, 2012), frame semantics (e.g., Fillmore & Baker, 2009), to cite a few. These theories differ in terms

of the types of information that enter into meaning representation, how features are combined, the nature of the primitive bases (viz., ontological categories upon which concepts are built), as well as the level, whether it be linguistic or conceptual, at which these representations are entertained.<sup>3</sup> But their commonalities, by far, surpass their differences, for they all seem to appeal to hidden predicates and other analytic properties to account for the semantic representation of lexical constituents and their carrier sentences.

We assume that semantic templates such as (3b) are intended to represent the *propositional* content of (3a) specifying its form and key elements of meaning.<sup>4</sup> The evidence corroborating this view either comes from distributional data or from experiments suggesting that complex templates are more difficult to process than simplex ones (i.e., they engender longer reading times; McKoon & Macfarland, 2000) or involve more “connections” (Gentner, 1981) between other simpler concepts in memory and are thus better recalled. We won’t repeat the review of the arguments and experimental studies supporting predicate decomposition, here (see de Almeida & Manouilidou, 2015; also Engelberg, 2011b): there seems to be widespread agreement of decompositional views, which spares us from a more thorough review. Our mission is rather to call attention to the evidence *against* decomposition, which also comes from distributional evidence and experiments—but which enjoy much less acceptance.

The first kind of evidence pertains to the lack of synonymy between sentences that are supposed to be semantically represented by the same constituents.<sup>5</sup> Take (4a) and (4b) as examples. These sentences, by hypothesis, yield the same semantic representation, as in (4c): while (4a) involves the lexical causative, (4b) involves its periphrastic counterpart. Unless the periphrastic *cause x to die* does not mean what is in (4c), the idea is that the two sentences are synonymous—hence that the template in (4c) should hold for *both* (4a) and (4b).

- (4) a. John killed the cat
- b. John caused the cat to die
- c. [[x ACT] CAUSE [BECOME [y <DEAD>]]]

But as Fodor (1970) argued sentences such as (4a) and (4b) do not denote the same events, for one can cause the cat to die on Saturday by poisoning his food on Thursday,

<sup>3</sup>We are assuming throughout that these theories all postulate that template structures are representations of psychological objects, as in Jackendoff (1983), similar to representations in a language of thought, though this is not always explicit in the works we cite.

<sup>4</sup>Although most of our discussion focuses on a theory such as Levin and Rappaport Hovav’s (2005), we assume that the main points we make apply to all theories we mentioned.

<sup>5</sup>An anonymous reviewer pointed out that, “Most people don’t assume that in order for there to be synonymy (and thus, analytic truths), the expressions in question need to be psychologically perfectly equivalent. For instance, it is standardly accepted that a correct analysis can be highly non-obvious.” We fail to understand what “most people” assume, for we do take synonymous sentences in natural language to be expressions of “perfectly equivalent” mental states (viz., propositions).

but one cannot kill the cat on Saturday by poisoning his food on Thursday. The distribution of time adverbials suggests that these are not similar events.<sup>6</sup>

Along similar lines, there are diverse experiments suggesting that causatives do not decompose, for they do not exhibit complexity effects (e.g., de Almeida, 1999a; Fodor et al., 1975, 1980; Kintsch, 1974; Manouilidou & de Almeida, 2013; Rayner & Duffy, 1986; Thorndyke, 1975; see de Almeida & Manouilidou, 2015, for review). These studies have employed numerous techniques—from judgment to reading times—and have been consistent in pointing to the lack of decomposition effects. More recently, data from Alzheimer's patients have also landed support to this camp. For instance, if verbs are represented by semantic templates, we should expect the pattern of deficits to reflect the purported effect of semantic complexity—with more complex concepts being harder to retrieve. Notice also in passing that the more predicates a template carries, the greater the chances that the concept might be impaired. But as we have recently shown (de Almeida, Mobayyen, Antal, Kehayia, Nair, & Schwartz, 2021), when Alzheimer's patients are asked to name video clips of events and states which depict classes of verbs with varying complexity (e.g., causatives, motion, and perception/psychological), these patients' naming pattern does not line up according to the predicted complexity. Causatives, which contain hypothetically more predicates are not affected as severely as psychological verbs, which contain less predicates. The pattern of results suggests that categorical deficits are not along the lines of semantic template complexity, but rather along the lines of thematic structure, with verbs assigning an *Experiencer* role to the subject position being harder to name. We assume that thematic roles are “psychologically real”: they affect the composition of a sentence in the mapping between syntax and the logical form, viz., by assigning roles to constituents based primarily on their syntactic positions and following the structural specifications of the predicate (see also Manouilidou, de Almeida, Nair, & Schwartz, 2009, for compatible results).

Crucially, the properties that enter into templates are far from well justified, for neither their ontological status has been determined, nor has the selection of features been principled.<sup>7</sup> At first, it may seem like a daunting task to think of a concept without thinking about the constituent parts we know (or more like *think*) to be true of that particular stimulus. For instance, it may be difficult to think of DRINK without entertaining thoughts such as LIQUID, or MOUTH. But entertaining these thoughts, as a function of entertaining DRINK does not necessarily entail that the likes of LIQUID and MOUTH are to be taken as *constituent features* of DRINK. Furthermore, if these features are taken to be constituents of DRINK, then, we can conclude that they too carry content themselves which are expressed in terms of

---

<sup>6</sup>This is perhaps old news but to our knowledge, with few exceptions (e.g., Jackendoff, 1990, 2002; Harley, 2012), it has not been addressed in the literature.

<sup>7</sup>As Jackendoff (2002, p. 377) puts it, lexical-semantic decomposition “... is a richly textured system whose subtleties we are only beginning to appreciate (...). It does remain to be seen whether all this richness eventually boils down to a system built from primitives, or if not, what alternative there may be.” While we take this position seriously, our point here is that the a/s distinction stands as the main obstacle to the empirical prospects of lexical semantics.

other features. The consequence of this is holism about content. And holism is the antithesis of semantics—as Quine had first suggested.

As a further example of this state of affairs, consider the distinction between so-called “externally caused” and “internally caused” change of state verbs such as those in (5a) and (5b) respectively.

- (5) a. The cement crumbled
- b. The apple rotted

Although much of this distinction bears on the realization of predicate-arguments (e.g., externally caused verbs usually do not enter into transitive forms), a critical issue is how the distinction is made in *semantic analysis*. For Levin and Hovav (1995), internally caused change of state verbs denote events brought about naturally in the object, while externally caused change of state verbs “imply the existence of an ‘external cause’ with immediate control over bringing about the eventuality described by the verb: an agent, an instrument, a natural force, or a circumstance” (p. 92).

The way the difference between these verb classes is presented appeals to our (perhaps naïve) knowledge of physics. But even that might fail us for we are not certain whether what makes something *rot* is internal or external, that is, whether atmospheric variables are the triggers of rotting, or alternatively if an object—say, an apple—rots entirely on its own. The same can be said of cement crumbling. The physics baggage is heavy. And we suspect this case lines up with classical cases of intentional fallacy plaguing semantics: even if the rot/crumble distinction can be determined solely on linguistic (viz., structural) principles, it is an entirely different claim to attribute the difference to mentally represented properties of the two types of events. Understanding the properties of the world will not help us fix the properties of semantic representations.

The point we are making, in summary, is one we have briefly touched upon in the previous section: just because one knows a stimulus or phenomenon to be composed of certain properties, it does not entail that these properties are encoded as *mental representations* of the stimulus or phenomenon. This is precisely the perennial effect of the intentional fallacy on semantic theorizing.

Before we further explore this issue, in contrast to atomism in Sect. 3, we would like to address rather briefly a second semantic phenomenon—*coercion*—one for which appeals to analyticity are also quite evident.

## 2.2 *Indeterminacy (or “Coercion”)*

The term “coercion” (or type-coercion, or type-shifting) is identified with particular hypotheses on how sentences such as (1) are interpreted—among which is the proposal presented in (2). We refer to these sentences as “indeterminate” because the actual action that Mary performed with the book is not determined, although the sentence is grammatical and a truth value judgment can be made (namely, it is true

if Mary began to do anything with the book); so much for terminology. The “coercion” hypothesis assumes that the proposition expressed by sentences such as (1) are necessarily enriched along the lines of what is exemplified in (2), but in particular proposal (2d), which we repeat here for convenience.

- (2) (d) Comprehenders incorporate the event sense into their semantic interpretation of the VP by reconfiguring the semantic representation of the complement, converting  $[\beta \text{ began}[\alpha \text{ the book}]]$  into  $[\beta \text{ began}[\alpha \text{ reading the book}]]$ . (Traxler et al., 2005, p. 5)

This processing hypothesis largely follows the theory of type coercion proposed by Pustejovsky (1995). The essence of coercion is that the alleged mismatch between the verb’s selectional restrictions and the nature of the internal argument. By assumption, the verb *begin* selects for an *event*, though the noun *book* is an *entity*. This mismatch triggers the search for a “plausible action” that would yield an enriched semantic composition, by interpolating a semantic constituent such as *reading* into the final form. But as we briefly alluded to in Sect. 1, a commitment to such a process entails a commitment to determining which, among all possible *senses*, are the ones to be interpolated into the resulting representation.

There is perhaps some confusion here between *meaning*, *sense*, and *use*—damage that unfortunately Wittgenstein cannot come back to repair. If we tell you that it is *hot* today, in Montreal, when actually it is  $-20^{\circ}\text{C}$ , we are most likely being sarcastic. It does not entail, now, that the concept HOT includes COLD, among its *senses*. We are certainly *using* the word *hot* to convey something else entirely, to provoke you or, as Davidson (1978) would say, to invite you to think, just like we would do with a metaphor. And even if we were to admit that *senses* are represented in close proximity (by some metric) with the original concept, as a function of extensive *use*, there is no saying on how a *sense* is to be accessed, other than via its actual host concept. Thus, to make a simple point: it is HOT that needs to be accessed such that COLD can be entertained.

It is clear that hypotheses committed to multiple layers of properties supposedly stored with token items are simply question begging: which sorts of elements are the ones to be chosen, and how are they to be chosen? As we will argue in Sect. 3, a different explanation can be offered in cases of conceptual tokening: inferences driven by synthetic relations are the ones that yield the *effects* which decompositionists claim to be effects of constituency. We will, thus, offer a more parsimonious analysis of this phenomenon, doing away with analyticity and placing the burden of interpretation on the identification of gaps, at the syntactic and logical-form representation of sentences, with most interpretation post-logical form being inferential, not relying on analytic properties of lexical concepts.

### 3 Alternative: Atomism and Inferences

What is, then, our proposal for doing away with analyticity? We should warn you that the proposal might be disappointingly simple, and our presentation of the theory will be somewhat constrained by the scope of the present chapter. Here is how we proceed. We start off by connecting our view of concepts with what we envision to be the architecture of cognition, as briefly presented in Sect. 1. Then, we discuss two main issues: (i) the representation of concepts according to our brand of atomism; and (ii) how concepts might be causally connected to each other—viz., as inferential relations. And, throughout, we tailor our discussion of atomism and inferences to the analysis of the two phenomena we discussed in Sect. 2.

We have mentioned that we are committed to symbolic representations and to computational processes. Patently, we take symbols that stand for content to be atomic, not molecular representations. And we take these symbols to compose into complex structures the classical way: complex symbolic expressions get their meaning as a function of the meaning of their constituent symbols and how they are arranged in propositions. Symbols then carry (or point to) information about the things (and events) they refer to. We do not establish a lower limit on the content that the simplex symbols convey—or more properly on the very content that they individuate—but we suggest that they are properties, predicates, and “particulars”, as Russell (1913) once put it. We assume that, for the most part, atoms are expressed by the simplex bound and free morphemes of natural language. And since we take concepts to be the very symbols of (again, Russell) our “experience”, we assume that they enter into different cognitive processes via computations.

So much for linking our view of conceptual representation and processes to the architecture we presented in Sect. 1. As for the nature of conceptual representation, if concepts are “atoms”, they are simply individuated by the kinds of things they refer. One quick note should suffice to address the problem of reference here: while we take concepts to be pointers to objects (in a very broad sense, including properties like patches of color) and events, they are also representations of things for which there is no referent (or, again, as Russell put it, in the “past, present, or not in time at all”, p. 5).

Two further observations are in order. The first is that it is likely that the things concepts individuate are full objects—the midsize things that populate scenes like chairs and pencils—or full events. But they can be just fractions of these: there is nothing in the system we suggest that ties the tokening of concepts to these ontological categories. And, to our knowledge, there is no clear line demarcating parts and objects, or objects and scenes (to wit, HORIZON is an “object” for all practical purposes; and so are DOG and TAIL). Second, a related issue: it is quite plausible to take “particulars” to be the tokening elements upon which one arrives at a given concept. For instance, it is well known that events have no fixed boundaries, that is, that the meaning of the verb *to kill*, say, does not pick up particular time and space properties, with well determined beginning and end points. Not even the property of being dead marks the endpoint of *kill*, for *to die* also lacks clearly perceptually

marked boundaries. Moreover, it is not the case that having *kill* entails having *dead*. In our system, the relation is inferential, not one of dependency.<sup>8</sup> If so, most likely the kinds of “particulars” that the conceptual system locks into may be the very entry points to the sets of inferences one runs in conceptual processing. This may become clearer with an example.

Take (6) to be the referential relation that obtains between the word (or the object) *dog* and its concept.

$$(6) \ dog \rightarrow \text{DOG}$$

The locking mechanism that affords DOG out of the word or object is a mechanism that in principle is tokened by whole objects, assuming that the visual attentional mechanism locks into full objects (see Fodor & Pylyshyn, 2015; Jackendoff, 2002). But it may well be the case that what one gets are *parts* of objects. Thus, getting TAIL tokened is what gets one to eventually entertain DOG. Notice that in order for this system to work, there ought to be a system of relations between concepts. As we mentioned above, we are committed to having conceptual relations that are *not* necessary; that is, to use the example, it is not the case that tokening TAIL necessarily causes DOG; only *tail* causes TAIL, but we suggest that one might get to the host object via its parts, not because they are conceptually dependent, but because they are inferentially connected.

We owe you, of course, a bit more clarity on how the system might work regarding these *non-analytic* inferences. We propose to work with the two phenomena we discussed in Sect. 2, beginning with causatives and, soon after, with the comprehension of indeterminate sentences. Along the way, we make a few observations regarding the less developed parts of our proposal.

### 3.1 Back to Causatives

Although we take Carnap’s commitment to analyticity in semantics to be misguided—just like Quine put it—the *tools* we inherited from him are of particular importance for conceiving psychological inferences bearing on meaning. Enter meaning postulates (henceforth MPs), which are quasi-logical inferences. We say quasi-logical only in the sense that they are not proper inferences whose consequent is *by necessity* entailed by the antecedent. And while this is a common tool in semantics, we take the kinds of MPs that run between concepts to be the very inferences

---

<sup>8</sup>We note in passing that, although this would take us far afield, what counts for us as a perceptual boundary for, say, *to die*, is tied to observation, not to the actual act of dying which is independent of observation. To wit, consider the end point of the verb *to break* as in *John broke the vase*: would it be when all physical particles of said vase cease moving? The concept BREAK is not determined by the actual physical phenomenon, by Newtonian laws (those are not “in the head”; cf. the intentional fallacy) but by when *break* causes BREAK.

that give rise to a myriad of relatedness *effects* found in the empirical literature and in other frameworks committed to analyticity.

Consider causatives. As we discussed above, voices in unison claim that causatives decompose. But there is strong evidence—from experiments and arguments—that causatives might not decompose. How, then, can one account for the pervasive effects obtained in the relations between arguments of the verb? How can one account for the pervasive effect of relations between transitive and intransitive variants of the same root verb? One way to conceive the relation between concepts—such that KILL and DIE or BOIL-transitive and BOIL-intransitive are related—could be by running inferences such as in (7).

- (7) a.  $\forall x \forall y [\text{BOIL}(x, y)] \rightarrow [\text{BOIL}(y)]$
- b.  $\forall y [\text{BOIL}(y)] \rightarrow [100^\circ\text{C}(y)]$
- c.  $(\forall x[P(x)] \rightarrow [Q(x)])_n$

We can cast this proposal in simple predicate logic, by attributing properties to individuals and by linking predicate relations as inferences. We can only highlight a few of the characteristics of this system—the ones that are in direct contrast with decompositional views discussed in Sect. 2. Notice also that the relation between transitive and intransitive variants of the same core concept can be accounted for by the entailment between arguments of the verb. But our suggestion is that beyond those entailments—which are in essence argument-structure driven—“properties” of the event denoted by the verb are also attained by these relations. We won’t extend this account of causatives here much further (but see de Almeida, 1999a, b, for early versions of this proposal). Suffice it to say that these inferences are not content-constitutive, thus, that it is not the case that the content of an utterance or a thought somehow depends on the “appropriate” inferences being computed. To us, the inferences that are typically run when concepts are tokened are *synthetic*, thus their actual content cannot be accounted for by semantic analysis.

We also acknowledge that even those with whom we share the main tenets of atomism have argued against adopting MPs for they are too unconstrained and thus cannot be used as an account of semantic inferences (Fodor, 1998). We part ways here. While we agree that they are unconstrained, our goal is not to model the very content tokened by a concept such as KILL or BOIL, but the inferences that might ensue that are taken to account for the conceptual content in all sorts of psychological effects (from priming to prototypicality to semantic-memory impairments). In summary, we suggest that inferences such as (7b) are entirely contingent on experience. And we suggest (7c) to be a basic law of how inferences run over predicates. To assume that those inferences constitute the representation of lexical content is, in principle to incur in the intentional fallacy.

### 3.2 Back to “Coercion”

We turn now to the other phenomenon, that of the comprehension of indeterminate sentences such as (1). To ease discussion and comparison with (2)—we will cast our proposal rather informally as in (8).

- (8) (a) Every incoming token lexical item (i) maps onto its corresponding concept (*book* → BOOK), (ii) contributes its syntactic information to the evolving syntactic tree (*book*  $\_N$ ), and (iii) contributes logical information to an evolving semantic composition (viz., a logical form;  $[\exists x, \text{BOOK}(x)]$ ).
- (b) The evolving syntactic parsing for a sentence such as (1) tags all its lexical constituents and its linguistically motivated gaps—viz., the gaps for syntactic positions that may be optionally filled-in lexically. As for (1), the gap is potentially in the VP, as in  $[\text{VP} [\text{V}^0 \text{ began } [\text{V}^0 e [\text{OBJ NP}]]]]$ .
- (c) The concepts that are accessed (mapped onto) by each lexical item are premises for *synthetic* inferences whose consequents are experience-based relations yielding between predicates (thus, a possible inference would be  $[\forall x \text{ BOOK}(x) \rightarrow [\text{READ[ABLE]}](x)]$ ).
- (d) The meaning of a sentence is obtained by combining the token concepts—the translations of morphemes—into the evolving logical form, such as  $\exists x (= \text{MAN}), \exists y (= \text{BOOK}) (\text{BEGIN } (x, y))$  (or, alternatively,  $\exists w (\text{BEGIN } (x, y, w))$ ); that is the shallow, *unenriched* interpretation of (1).
- (e) Many processes of enrichment ensue; among them are the processes of filling the gaps identified during syntactic structuring with the concepts that were part of the postulates triggered by (i) the utterance context, and (ii) the co-text.

We can only make brief observations about (8)—but we trust that the contrast with (2) is quite clear. First, notice that the *meaning* of *book* is not a *sense*; and, according to our proposal, there are no senses stored *with* the meanings of words. We do not deny that there are *uses*, but *uses* are obtained pragmatically (they are *synthetic*; see below), within the inferences that run after conceptual tokening (as in 8a) and conceptual composition. Also, as suggested in (8b) there are linguistic arguments for holding a syntactic gap within the VP of sentences such as (1) without appealing to effects of “coercion”.<sup>9</sup> And we hold that the coercion effects shown in most experimental studies could be effects of this gap as they can also be effects of inferences that the indeterminate sentence triggers.

The advantage of a proposal such as the one sketched in (8), in summary, is that it does away with analyticity. For any of the proposals appealing to analytic properties, the burden is to determine the criterion for separating analytic from synthetic

---

<sup>9</sup> Several linguistic arguments for the VP gap hypothesis appear in de Almeida and Dwivedi (2008) and in de Almeida and Riven (2012). Also, see arguments against coercion alternatives in de Almeida and Lepore (2018) and in de Almeida et al. (2016), which we cannot begin to discuss here.

properties. We do not appeal to such properties because to us concepts are atomic, but we see a role for such properties in the *inferences* that ensue upon conceptual tokening and semantic composition.

### 3.3 Conclusion: Atomic Concepts and Inferences

We conclude by stressing a few points about our proposal. First, in the sense we take in the present proposal, the inferences about lexical-conceptual properties are mostly (if not all) synthetic, not analytic, as mentioned above. Thus, one can know what a dog is without knowing what an animal is or what a pet is, for that matter. Crucial to this approach is the idea that all such relations, commonly known as constituent *features*, are synthetic and thus the inferences that run over them are not *necessary* for content attainment. In fact, only the content that each individual symbol instantiates suffices, independent of the inferences it generates. If inferences are synthetic, they cannot be part of the meaning of a token item. And if they are not part of meaning, we can dispense with a semantics that attempts to legislate on experience and world knowledge.

Second, we assume that many of the inferences that run as a consequence of a concept being triggered are common to many inhabitants of the same community, those sharing similar kinds of experiences. We cannot be precise on this idea because it points to something whose variables are virtually infinite. Crucial to our approach, in fact, is the idea that these commonalities cannot be legislated on. We also suggest that many, perhaps most effects found in the literature—from priming to prototypicality—are manifestations of these inferences; they are effects of the causal connectedness established between concepts as a function of use and experience. And we even acknowledge that it may be difficult to dissociate—empirically—between inferences computed upon tokening concepts and effects of “activation” of properties. However, we have presented some clear signs from the literature that point against decomposition.

We do hold that there is a crucial distinction, upon which a theoretical advantage stands: by not taking properties to be analytic, there is no commitment to building a semantic theory whose foundations are faulty. The crucial distinction between atomism and molecularism is that the former, but not the latter does not require semantic analysis based on features or synonymy and, because of that, there is no analysis of content other than assuming that concepts (and their lexical labels) are largely referential, symbols that point to things, events, ideas, and so forth. Reference does not entail being in the presence of the object or event: it entails bringing to fore the relation between the symbol and the thing/event/idea it designates.<sup>10</sup>

If semantics appeals to features, without an analytic/synthetic distinction, it turns to holism, which is the antithesis of semantics—at least of a semantics committed to

---

<sup>10</sup>This point was made by Russell (1913, Chap. 3) and, more recently, by Fodor and Pylyshyn (2015, Chap. 5) regarding reference “beyond the perceptual circle”.

compositionality and productivity. If semantics appeals to properties of the world to fix properties of mental representations, it may fall into the intentional fallacy trap. The way semantics can avoid all this trouble is to turn to atomism cum inferences.

**Acknowledgements** We are grateful to two anonymous reviewers for very useful comments. We are aware we did not address all their concerns, but we hope to have further clarified our position. We also thank the National Sciences and Engineering Research Council of Canada (NSERC) for financial support.

## References

- Carnap, R. (1956). *Meaning and necessity* (2nd ed.). Chicago University Press.
- Croft, W. (2012). *Verbs: Aspect and causal structure*. Oxford, UK: Oxford University Press.
- Culicover, P. (1970). One more can of beer. *Linguistic Inquiry*, 1(3), 366–369.
- Davidson, D. (1978). What metaphors mean. *Critical Inquiry*, 5(1), 31–47.
- de Almeida, R. G. (1999a). *The representation of lexical concepts: A psycholinguistic inquiry*. Unpublished doctoral dissertation, Rutgers University, New Brunswick, NJ.
- de Almeida, R. G. (1999b). What do category-specific semantic deficits tell us about the representation of lexical concepts. *Brain and Language*, 68, 241–248.
- de Almeida, R. G. (2018). Composing meaning and thinking. In G. Preyer (Ed.), *Beyond semantics and pragmatics: New essays* (pp. 201–229). Oxford, UK: Oxford University Press.
- de Almeida, R. G., & Dwivedi, V. (2008). Coercion without lexical decomposition: Type-shifting effects revisited. *Canadian Journal of Linguistics*, 53(2), 301–326.
- de Almeida, R. G., & Lepore, E. (2018). Semantics for a module. In R. G. de Almeida & L. Gleitman (Eds.), *On modules, concepts, and language: Cognitive science at its core*. Oxford, UK: Oxford University Press.
- de Almeida, R. G., & Manouilidou, C. (2015). The study of verbs in cognitive science. In R. G. de Almeida & C. Manouilidou (Eds.), *Cognitive science perspectives on verb representation and processing* (pp. 3–39). New York, NY: Springer.
- de Almeida, R. G., Riven, L., Manouilidou, C., Lungu, O., Dwivedi, V. D., Jarema, G., et al. (2016). The neuronal correlate of indeterminate sentence in interpretation: An fMRI study. *Frontiers in Human Neuroscience*, 10, 614. <https://doi.org/10.3389/fnhum.2016.00614>.
- de Almeida, R. G., Mobayyen, F., Kehayia, E., Antal, C., Nair, V., & Schwartz, G. (2021). Category-specific semantic deficits for verbs in Alzheimer's disease: Evidence from static and dynamic action-naming tasks.
- Engelberg, S. (2011a). Frameworks of lexical decomposition of verbs. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (Vol. 1, pp. 358–399). Amsterdam: De Gruyter Mouton.
- Engelberg, S. (2011b). Lexical decomposition: Foundational issues. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (Vol. 1, pp. 124–144). Amsterdam: De Gruyter Mouton.
- Fillmore, C. J., & Baker, C. (2009). A Frames approach to semantic analysis. In B. Heine & H. Narrog (Eds.), *The oxford handbook of linguistic analysis*. Oxford, UK: Oxford University Press.
- Fodor, J. A. (1970). Three reasons for not deriving 'kill' from 'cause to die'. *Linguistic Inquiry*, 1, 429–438.
- Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford, UK: Oxford University Press.
- Fodor, J. A., Garrett, M. F., Walker, E. C., & Parkes, C. H. (1980). Against definitions. *Cognition*, 8(3), 263–367.

- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1), 3–71.
- Fodor, J. A., & Pylyshyn, Z. W. (2015). *Minds without meanings: An essay on the content of concepts*. Cambridge, MA: MIT Press.
- Fodor, J. D., Fodor, J. A., & Garrett, M. F. (1975). The psychological unreality of semantic representations. *Linguistic Inquiry*, 6, 515–531.
- Gentner, D. (1981). Verb semantic structures in memory for sentences: Evidence for componential representation. *Cognitive Psychology*, 13, 56–83.
- Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: Cambridge University Press.
- Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.
- Jackendoff, R. (2002). *Foundations of language*. Cambridge, MA: MIT Press.
- Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale: Lawrence Erlbaum Associates.
- Levin, B., & Hovav, M. R. (1995). *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge, MA: MIT press.
- Levin, B., & Rappaport-Hovav, M. (2005). *Argument realization*. Cambridge, MA: Cambridge University Press.
- Manouilidou, C., & de Almeida, R. G. (2013). Processing correlates of verb typology: Investigating internal structure and argument realization. *Linguistics*, 51, 767–792.
- Manouilidou, C., de Almeida, R. G., Schwartz, G., & Nair, N. P. V. (2009). Thematic roles in Alzheimer's disease: Hierarchy violations in psychological predicates. *Journal of Neurolinguistics*, 22(2), 167–186.
- McKoon, G., & McFarland, T. (2000). Externally and internally caused change of state verbs. *Language*, 76, 833–858.
- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99, 440–466.
- Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.
- Quine, W. V. O. (1953). Two dogmas of empiricism. In *From a logical point of view* (2nd ed.). Cambridge, MA: Harvard University Press.
- Rayner, K., & Duffy, S. (1986). Lexical complexity and fixation times in reading: effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition*, 14(3), 191–201.
- Russell, B. (1913/1984). *Theory of knowledge: The 1913 manuscript*. London: Routledge.
- Titchener, E. B. (1909). *Lectures on the experimental psychology of thought-processes*. New York, NY: Macmillan.
- Traxler, M. J., McElree, B., Williams, R. S., & Pickering, M. J. (2005). Context effects in coercion: Evidence from eye movements. *Journal of Memory and Language*, 53(1), 1–25.
- Thorndyke, P. W. (1975). Conceptual complexity and imagery in comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, 14, 359–369.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Linguistic Relativity and Flexibility of Mental Representations: Color Terms in a Frame Based Analysis



Leda Berio

**Abstract** This paper connects the issue of the influence of language on conceptual representations, known as Linguistic Relativity, with some issues pertaining to concepts' structure and retrieval. In what follows, I present a model of the relation between linguistic information and perceptual information in concepts using frames as a format of mental representation, and argue that this model not only accommodates the empirical evidence presented by the linguistic relativity debate, but also sheds some light on unanswered questions regarding conceptual representations' structure. A fundamental assumption is that mental representations can be conceptualised as complex functional structures whose components can be dynamically and flexibly recruited depending on the tasks at hand; the components include linguistic and non-linguistic elements. This kind of model allows for the representation of the interaction between linguistic and perceptual information and accounts for the variable influence that color labels have on non-linguistic tasks. The paper provides some example of strategy shifting and flexible recruitment of linguistic information available in the literature and explains them using frames.

**Keywords** Colors · Labels · Concepts · Perceptual information · Frames

## 1 Introduction

Cross linguistic<sup>1</sup> research about basic color terms has been for a long time a central concern in the debate regarding Linguistic Relativity, i.e. the influence of language on conceptual representations. However, this has been seldomly connected to the

---

<sup>1</sup>Many thanks are owed to Kurt Erbach, Gottfried Vosgerau and the Ph.D. students of the SFB 991 in Düsseldorf for discussing various iterations of this work, to Alexandra Redmann and Natalja Beckmann for discussing the frames, and to my reviewers. This research was funded by the German Research Foundation (DFG) CRC 991, Project D02.

---

L. Berio (✉)

Department of Philosophy, Heinrich–Heine–Universität, 40204 Düsseldorf, Germany

e-mail: [leda.berio@hhu.de](mailto:leda.berio@hhu.de)

issue of the *structure* of mental representations. In this paper, I will argue that a frame-based model of mental representations allows for the representation of the relation between the perceptual information contained in color concepts and their linguistic labels in a way that is compatible with the empirical evidence used in the Linguistic Relativity debate. In doing so, I shift the problem of Linguistic Relativity to a matter of the structure of mental representations. In the account I present, mental representations are conceived as complex functional structures that are dynamically and flexibly recruited according to the task at hand and that include both linguistic and non-linguistic information. The core claim of the paper will then be that such a model allows for the presentation of the interaction between different components of a mental representation and can account for the variable influence of linguistic labels on color-related tasks in terms of strategy shifting and flexible use of mental representations' components.

In the first part of the paper, I delineate the debate about Whorfianism and its more recent declinations, connecting the debate to the problem of flexibility in mental representations. Secondly, I briefly present a few examples of effects of what is called “shallow Whorfianism”, describing the available experimental evidence. In the third section, I propose a way to represent color concepts in frames and I subsequently show how this can be applied to concepts in general. In Sect. 4 of the paper, I explain how this view can be fruitfully applied to communicative situations and pragmatic effects and, most importantly, to model the experimental data presented in Sect. 2. In Sect. 5, I provide an example from a different conceptual domain (number representation) that can be treated efficiently with the proposed model. In Sect. 6, I show how, in the same spirit, the model can be used to model a classical color task, i.e. the Stroop task. Finally, I draw conclusions regarding the debate and suggest further necessary steps.

## 2 Color Terms and Whorfianism: Some Coordinates

### 2.1 *Universalism, “deep” and “shallow” Whorfianism; Intertwined Issues*

For a long time, the debate regarding color terms acquisition has been influenced by a (sometimes well grounded) bias against the idea of Linguistic Relativity: one of its earliest formulations, namely the Sapir-Whorf hypothesis, suggests as a matter of fact a particularly strong and simplistic influence of language on thought. However, the debate has seen a partial re-ignition due to more modern studies and techniques that, revisiting the Whorfian hypothesis' too strong initial assumptions and statements, have postulated a role for language in various tasks. This is also partially due to the fact that what was initially taken as the final word on the color terms debate (namely the study by Berlin and Kay 1969) has been scaled down to be an important but not decisive piece of evidence. This is not the place to discuss Berlin and Kay's

research and proposal for universal patterns in color terms; for the present purpose, it is sufficient to keep in mind that it is possible to postulate some kind of influence of color terms on color cognition without necessarily contradicting Berlin and Kay's fundamental insight that there are universal tendencies and/or constraints on focal colors that are perceptually more salient and therefore easier to identify in absence of corresponding color terms.

It is essential to specify that this debate is concerned with a particular aspect of language, which is indeed lexical labeling: most studies regarding color cognition are focused on whether or not color terms that are present in one language have any influence on performance as far as color recognition is concerned. This brings us to the other important specification, which is that the debate is concerned with influence on perception and categorization tasks. The color words debate is often enough considered the privileged (if not exclusive) ground for deciding about the whole debate concerning Whorfianism and Linguistic Relativity. However, it is worth underlining that the main focus of a big part of the debate is very specific: whether or not lexical entries influence perception and attention mechanisms.

As a matter of fact, as Lalumera (2014) already notices and as it will be clear in the next paragraphs, the evidence available in the literature cross-cuts the distinction between Whorfianism and Universalism, since there are in this sense various kinds of results suggesting, on the one hand, some influence of linguistic labels on perception mechanisms, and on the other hand, rejecting the extreme claim made by language relativity supporters in the past, namely that language strongly shapes mental representations. Thus, the distinction between Universalism and Language Relativism has partially been replaced in the literature by what Lalumera phrases as a distinction between "deep" and "shallow" Whorfianism, separating those phenomena where the influence of linguistic labels seems to be constant, pervasive and stable, from those cases in which it is "only" a flexible, context dependent, task dependent influence of some sort. The reason why this distinction cross-cuts the previous one, i.e. Universalism vs. Whorfianism, is that the old debate was concerned with a less fine-grained question: through the universalist lenses, Whorfianism was seen as threatening the idea of concepts as something that follows potentially the same "rules" of formation and development regardless of the language of the speaker, therefore menacing the idea that humans have a somehow universal conceptual repertoire. Whorfianism, on the other hand, was concerned with the fact that universalism seemed not to admit any interference of language with mental representations' structure and complexity. Framing the debate as "deep" and "shallow" Whorfianism shifts the focus of the debate to a somehow more pragmatic issue, namely how do linguistic processing and linguistic labeling interfere with non-linguistic processes, including but not confined to conceptual formation, and to what extent is that relevant in non linguistic tasks. The question then becomes, when is this influence relevant and how stable and pervasive is it. In what follows, I will also try to argue that this might shed some light on how to think of conceptual structure itself, without making the bold, original Whorfian claim that language invariably shapes representations.

Note that this whole debate is better understood if connected with the parallel but distinct issue regarding cognitive penetrability.<sup>2</sup> Cognitive penetrability can be defined as the property of perceptual experience to be influenced by what happens at the so-called higher cognitive level; in other words, we speak of cognitive penetration when perceptual experience is influenced by beliefs, desires, intentions and concepts (Newen and Vetter 2017). In a way, the debate can be conceived to proceed hand in hand with the issue treated here: admitting an influence of linguistic information on non linguistic processing means admitting permeability of perceptual experience. The problem of permeability, on the other hand, is of a broader nature, as it comprises considerations regarding modularity and specialization of brain areas; in other terms, the debate regarding permeability brings us to a broader scale of issues regarding cognition in general. The focus of the current paper is on the relation between linguistic labels and color *concepts*; which means, on the one hand, that perception is obviously relevant for the discussion, given color perception is at the center of the debate; but also, on the other hand, that the focus is already on mental representations employed in experience and not on perceptual experience itself, which implies that the focus is on the level of “higher cognition” only.

Admitting permeability means admitting that the *experience* of color changes depending on (among other things) linguistic processes; the debate regarding Linguistic Relativity focuses on whether or not the *concepts* related to color and used in perception are influenced by color labels. This claim is therefore both weaker and related. Related, because color mental representations are supposedly recalled in color perception; but weaker, because it moves prevalently at the level of higher cognition (linguistic information influencing representations) and because it does not make claims on the *experience* related to color but only on the representational means employed.<sup>3</sup>

As it will be clear in the rest of the paper, the view proposed here, despite being mainly concerned with mental representations and higher cognition as said, assumes permeability. As a matter of fact, it is assumed here that different kinds of information such as perceptual and motor information are integrated in mental representations along with more abstract kinds of information, like linguistic-based one. In this sense, the view even *endorses* an account of mental representations that accepts cognitive penetration and refuses strict modularity.

Getting back on the shallow–deep spectrum, “deep Whorfianism” is problematic to argue for, given the scarce evidence in favour of an influence of language on thought that actually is not task dependent but stable and pervasive. Moreover, it is arguably a type of influence that is more likely to be related to words and concepts that are more complex and less perceptually-bound than color ones, as it will be argued elsewhere.<sup>4</sup> However, the focus of this paper is the so-called “shallow” Whorfianism, or, in other

<sup>2</sup>Thanks to the anonymous reviewer for pointing out the necessity of mentioning this.

<sup>3</sup>Note that Macpherson (2012) contains an interesting review of color literature connected to cognitive penetration.

<sup>4</sup>One assumption of my work on the interface between language and cognition is that it varies depending on the type of concept/category that is considered.

words, the influence of language that is only detectable in specific tasks. In the frame of the Universalism-Whorfianism debate, this kind of influence is irrelevant, because the question at issue is whether having a different language irreversibly shapes the conceptual repertoire in a deep, pervasive way. In this sense, the answer going along with shallow Whorfianism is, clearly, negative. However as Lalumera points out.

[...] some Whorfian effects show themselves to be task dependent and temporary. A question on this point is worth raising here. Is that enough to deem such effects as uninteresting, qua task dependent and temporary? The answer is that it would be enough, but at the price of committing to the view that only stable and context-free representations are employed in perception and cognition. (p. 7).

This is an essential remark: arguing against any kind of influence of language on non-linguistic cognitive processes appealing to the fact that the supposed influence might only be task dependent and not *always* present means endorsing a view of mental representations that is not trivial (anymore). In other words, it means committing not only to the idea that there is a stability in mental representations and categories, but also that this stability is such that everything that regards the flexible, online, task dependent application of these same categories is not relevant because it does not tell us anything about mental processes. Lalumera points out that this does not seem to be the case, and that there is plenty of evidence suggesting the contrary. My claim goes in a slightly different direction: I think that what the evidence available in the literature suggests is that a way to represent the interaction between linguistic labels and conceptual units is needed and that, whatever the model, it has to cope with how variable this influence actually is. In what follows, I will briefly present some examples of “shallow Whorfianism” that are present in the literature and then propose a way to model them using frames. I will then try to show how the model can be flexible and fruitful in dealing with some challenges that conceptual representations and language present to us, if we assume a view of representations as flexible adaptable structures that can be differentially activated depending on the task at hand.

## 2.2 “Shallow” Effects of Color labelling

Many examples in language cognition and color deal with perception tasks. In this paragraph, I will focus on two well-known studies that are often referred to in the literature because they’re considered evidence that Whorfian influence is “shallow” because it is task dependent. Later in this paper, I will focus on one of them as a paradigmatic case that points in the direction of a flexible, context dependent use of linguistic representations in non-linguistic tasks, while at the same time underlining the open questions that are left.

A well known and cited study, therefore worth mentioning as a valid example, is Winawer et al. (2006). Russian has an obligatory distinction between *light blue* and *dark blue* (*goluboy* and *siniy*), as many other languages, like Greek and Italian, do.

In the study, subjects (divided between Russian speakers and English speakers) were shown three color squares arranged in a triad; the task consisted of saying which one of the bottom squares was identical to the one on top, while reaction times were measured. In “within category” trials, the square was from the same color category of the match, whereas in “cross-category” trials the distracter and the match belonged to different categories in Russian color categorization system.

The hypothesis was that the presence of a color boundary available in one language (Russian) but not the other (English) would have affected performance across the boundary; more specifically, that Russian speakers would have made faster cross-category discriminations than within category ones. The prediction was confirmed: there was indeed a difference between the performance of Russian speakers and that of English speakers. Even more interestingly, the effect disappeared if the subjects also had to perform a verbal interference task at the same time (the task consisted in silently rehearsing digit strings): it seemed, then, that blocking language resources with task-irrelevant processing was preventing the effect. At the same time, estimating the difficulty of the trials, the research group found out that the difference between cross-category and within-category trials performance for Russian speakers increased the more difficult the discrimination was.

Several interpretations can be given of the results. First of all, the fact that the facilitation disappears when linguistic interference is added, suggests at least two things: firstly, that the effect on perception is temporary and tied to the specificity of the task, and secondly, that language labels are extremely likely to be the cause of the effect, because linguistic coding seems to be involved. Clearly, then, we are in the realm of what has been referred to as “language as a meddler” (Wolff and Holmes 2010): there is an *online* interference that takes place during a certain task and that is heavily dependent on the context and conditions of the task itself. It is also clearly a case of language *changing* the performance as far as an already existing skill is concerned, namely, to be precise, color discrimination. One of the most interesting results is definitely that the difference in performance increased if the task was perceptually more difficult: this suggests that language was used as a facilitator of some kind, with linguistic labels possibly used too, as a support for the difficult discrimination task. In this case, then, we have a case in which language is *improving* the performance on a task.

Different kind of data comes from studies like that of Roberson et al. (2008), who explored differences between English and Korean speakers. Korean has fifteen basic color terms, as opposed to the eleven English ones. Once again, color perception was the focus of the study, which was aimed at comparing linguistic distinguishability and perceptual one. It is often argued that language centres are to be located on the left hemisphere and categorization functions are to be attributed to clusters in the right hemisphere; wanting to test this distinction, the study investigated the categories of *yeoundu* and *chorok*, respectively *yellow-green* and *green* in Korean. In the task, participants were presented with an array of color patches, among which one was different from the others. The patches all belonged to the category *green* for English speakers; for Korean speakers, however, the “odd ball” patch could belong either to the same category as the others or not. Participants had to say whether the odd ball

was right or left in the screen (hence, the stimulus was presented to be elaborated either in the right or in the left hemisphere). Once again, there was a difference in cross-category and within-category discrimination: Korean speakers made faster cross-category judgments compared to within category ones; the effect was present regardless of the visual field. However, a comparison between fast responders and slow responders led to an interesting result; fast responders only were facilitated when the stimulus was presented in the right visual field, whereas the effect was present for slow responders even for the left visual field-presented stimuli. This was interpreted as a sign that the effect was due to linguistic labels: in case of slower responses, time allowed the information to be transmitted via *corpus callosum*. Even here, the influence of language labels is evident, but at the same time clearly dependent on task constraints. Similarly to the previous case, moreover, we are talking about an influence of language labels on perception and attention mechanisms.

In both the mentioned cases, there is an influence of language that is clearly constrained by determined conditions and tasks: moreover, these are not isolated cases. Evidence very similar to Roberson et al., for instance, was collected by Gilbert and colleagues (2007). In general, what this kind of evidence tends to suggest is that influence of color words is variable and task dependent, and this seems to be suggested by other studies as well in other semantic domains (see Papafragou, 2008 for instance). However, these results, while suggesting cognitive penetration of some kind, still do not shed any light on what the possible relation between linguistic labels and mental representations is and how it can be modeled.

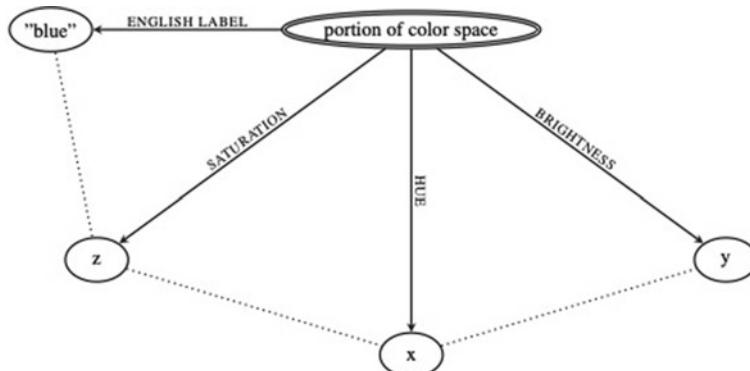
### 3 Frames and Representation of Colors

Let us take a step back and consider the kind of picture that is compatible with the presented data. As underlined, this kind of data is often cited in the domain of Linguistic Relativity as an influence of language on color concepts; however, little is said about how color concepts enter the picture.

There are several accounts out there that try to tackle the issue of the structure of mental representations, and this paper is not meant to be a review of them; on the other hand, it is at least worth underlining that papers as influential as the one published by Casasanto and Lupyan (2019) efficiently sum up plenty of good evidence in favour of representations as task and context dependent in various ways, showing how evidence from psycholinguistic and cognitive science accounts for a great flexibility in mental representations.<sup>5</sup> In what follows, I will adopt the idea that concepts can be efficiently represented as frames as developed by Barsalou (1992). There exist several theoretical elaborations of frame theory and the research regarding its compatibility with other theories of mental representation is vast; for the purpose of the paper,

---

<sup>5</sup>Casasanto and Lupyan use this evidence to argue, at the same time, against (1) the idea that there is any stability in mental representations (2) the possibility of talking about *shared* representations. I think their claim is, in this sense, far-fetched, but this goes outside the scope of this paper.



**Fig. 1** Frame for the color concept BLUE

however, only a few specifications are needed, starting from the idea that frame theories assume that an efficient way to describe and model conceptual components is to think of complex structures where attributes get assigned unique values.

Furthermore, note that frame theories are quite different from feature lists approaches, for instance, or from concept atomism, since they all assume that concepts have a fine-grained complex structure (contra atomism) and that attributes are functional, contra feature list approaches.<sup>6</sup> However, choosing frames as a model, in this instance, does not mean necessarily buying one specific philosophical theory of concepts. Assuming this is a good model for conceptual representations does not mean necessarily take a stance on the issue, for instance, of whether or not prototype theory is a good account for concepts; there is currently a lot of research regarding how and when frame theory can be integrated in other approaches, and that heavily depends on the kind of frame theory that is chosen. For the purpose of this paper, however, only two characteristics of frame theory have to be assumed: the possibility of building recursive structures (1) and the possibility of imposing *functional* relations and constraints among attributes and nodes (2).

Let us assume that labels for colors can be considered as an attribute, *label*, functionally connected to another node in an attribute-value structure.<sup>7</sup>

The frame for a color concept then would look like Fig. 1. The expression “portion of color space” is here intended as a place holder for a region of the color space, i.e. a value interval (note that thinking about it in terms of a prototypical blue or an exemplar-like blue does not make a difference for the present purpose). The arrows in the frame represent the functional attributes; the non-arrow arches represent constraints between the attributes. Roughly speaking, the idea is that a color concept can be represented in terms of a portion of color space characterized by a given

<sup>6</sup>This is a characteristic of Düsseldorf frame theory, adopted in this paper; see Löbner 2015.

<sup>7</sup>Modelling the relation between linguistic information and conceptual one, far from being contradicting frame theory, is also the focus of other current research. For a compatible account see for instance Beckmann, Petersen and Indefrey, submitted).

*saturation*, *hue* and *brightness*, whose value range constraints the attribute *English label*. Ideally the constraint can be spelled out in these terms:

$$\text{If}(x \in \{..\}, y \in \{..\}, z \in \{..\}, )\text{then } \iota = \text{"blue"} \quad (1)$$

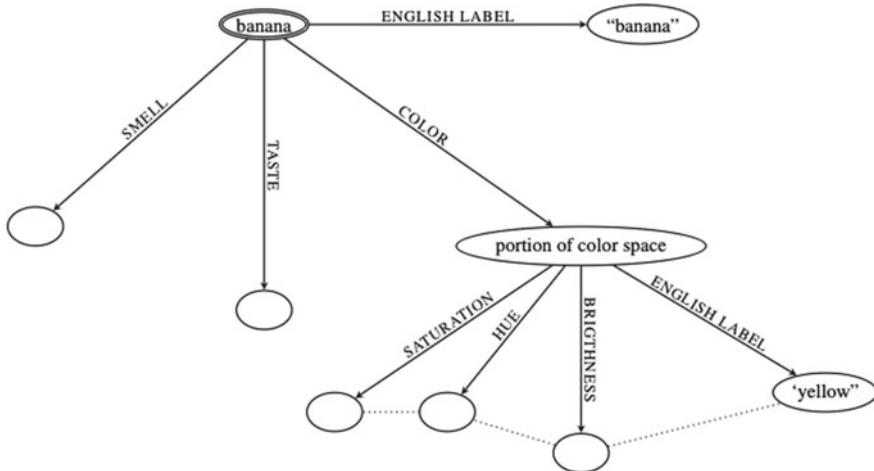
where  $\iota$  represents the value of the attribute *English label*, which is in this case “blue”. The formula reads so that, if the values of *hue*, *brightness* and SATURATION are included in a given interval, then a given label applies to the portion of color space considered.

Note that there is a clear difference between attributes like *hue*, *brightness* and *saturation* and one like *English label*. In the first case, we have information whose knowledge does not have to be declarative, whereas in the latter we have a linguistic attribute of which we necessarily have a declarative knowledge. This is not problematic because the frame does not represent the declarative knowledge about a color, but rather the structure of the representation. This applies even more significantly to the values that the attributes take, since it might be explicit in my representation that colors are characterized by these three aspects, but I might not know the values involved. Clearly, the idea for these three attributes is that the values they take range in a determined interval. The importance of specifying the language considered should be clear; the idea is that different languages will have different constraints operating (constraints where the intervals for the values of hue, brightness and saturation are different) and will give different results in terms of the label. Another obvious necessity of specifying the language in the attribute will be, for instance, considering the fact that bilingual speakers might have more than one label available for the same values  $x$ ,  $y$  and  $z$ . Such a mental representation, then, contains both explicitly known and implicitly known information, represented by values that can be either an interval or not, depending on the kind of attribute.

Let us embed a frame for a color concept like this one in a different frame, in Fig. 2. The given example illustrates a frame for the mental representation of a banana. Clearly much more than what is represented could enter a speaker’s representation of a banana, but only salient or situationally-relevant attributes are listed in the representation. The underlying idea is that this might be a way to represent what an individual speaker has in mind when thinking about a banana.<sup>8</sup> Clearly, an assumption here is that the linguistic label for an object, like for instance a banana, is part of the set of information connected to the perception of the object in the mind of the speaker or, in other words, that it makes sense to think about the semantics of word meaning not to be disconnected from mental representations of the objects that words denote. The advantage of such a move will hopefully be clear once we will be proceeding with the rest of the argument.

---

<sup>8</sup>Albeit, again, with all the simplifications applied here for the sake of brevity. The individual’s representation of a Banana might include a lot of idiosyncratic information: judgements about how bananas taste like, for instance, or individual experiences concerning this type of fruit, or even some kind of danger signal in case of an allergy to bananas. The amount of idiosyncratic information included in a frame is a matter of discussion.



**Fig. 2** Instantiated frame for a banana

First thing to notice is that the frame includes information that is basically only perceptual in one of the nodes.

The idea is that a flexible structure like a frame (or, better, the interaction between frames) can be used to incorporate different sources and kinds of information, including purely perceptual one. The intuition under this frame is that different essential features of “banana” are listed that constitute some of the relevant parts included in an individual’s representation of what a banana is. Other standard attributes we probably might associate with it include, for instance, SHAPE. COLOR is also a standard attribute; what is fundamental here is that frames are recursive, combinable structures. In this case, the color of a particular banana the speaker might have in mind is related to the concept of that color, which might be an exemplar-like representation or a prototype, for example. This concept is then labeled in English. Just like in the “banana” case, the label is considered an attribute among others in the mental representation. The suggestion, then, is to consider the fact that an attribute like *English label* can be inserted and that it applies to both the color and other features of the frame.

Note, furthermore, that the frame represents the banana in the context of ripeness; it is clear that in another context the value for the functional attribute COLOR could be a different portion of the color space (since, for instance, we would have a brownish color when seeing a overripe banana, or a greenish color when seeing one that it’s not ripe enough). In that case, the values for the attributes *saturation*, *brightness* and *hue* will be different, and depending on the constraints operating on the language, the resulting label will be different.

Now, one of the advantages of frames is that they spell out the functional relationships between elements of the representations and, therefore, can be used to give a

picture of what happens during communication in an effective way. In the next session, I will briefly discuss two kinds of communicative phenomena that can involve color words.

## 4 Color Words and Flexible Use of Representations' Features

A characteristic of communication involving color words is that it can give rises to interesting phenomena; to proceed with the argument, let us consider some of the most common examples that can be given when treating the sorites paradox or models of vagueness (see for this variant Rayo 2011). Having a grayish-blueish house among a group of houses that are painted in red and green, we can successfully utter.

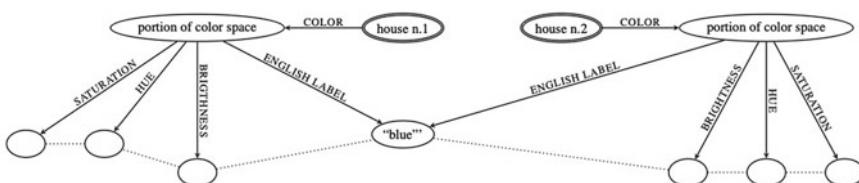
[1] Peter's house is the blue one.

and be understood as indicating the grayish-blueish house. In this context, the portion of color space the color of the house can be placed in can be labeled correctly.

However, in a context where the block consists of a blue house, the same blueish-grayish house, a red house and a green house, [1] cannot be used to point to the second one. In this case, "blue" does not apply correctly (or, at least, it does not represent the most successful communicative choice), even if we are considering the same portion of perceptual space. In other words, the label we are using in communication has to change to make the conversational exchange effective. The value of the attribute, then, will vary.

Integrating the two frames representing the two houses can help (Fig. 3); the strategy of labeling the grayish house (house number 2, for instance), "blue" is not a felicitous one because it means recalling the same label used for house number 1; given that the task includes differentiating between the two houses, having the same label does not aid the discrimination and it's therefore not a winning strategy, communicatively speaking. In this context, the discrimination task cannot succeed because the label can be applied to both houses. The frame representation makes the pragmatic effects, in this way, very easy to spot.

The first type of variability I want to draw attention to is therefore this one; color labels for the same portion of color space referring to the color property of an



**Fig. 3** Two houses' frames

object vary in their communicative efficacy. It is essential to stress that this is a point regarding how mental representations are *used* in *communication*. It is certainly true that, giving an array of color terms available and wanting to apply them in a rigorous way to a representation of color space, we do not have the same kind of phenomenon, but rather a series of determinable-determinate relations: hence, a portion of color space “blue” that can be labeled, on a more fine grained level, “ultramarine” and another that can be “Nivea blue”.<sup>9</sup> However, what is meant with the given example is something different, i.e. that a communicative situation can make a label for a determined color more or less communicatively efficient and appropriate in a context, even more so in Sorites-like cases, where this depends on whether or not the perceived color is close in perception to other present portions of the color space. Frames make it particularly easy to see, granting a format of mental representation modeling that aids the understanding of pragmatic effects.

There is also another element of variability, namely the relevance that the activation of a determinate attribute (and therefore of the respective value) has in a determined situation. In other words, at least as far as a certain understanding of frame theory is involved; attributes can be activated or not during tasks that involve the representation in question. Let me use another example at the intuitive level to express the idea. Let us assume I ask a colleague to hand me a folder in my office that contains the notes from the Dynamic Semantics class I am following. The colleague knows me and my office and knows that my folders are all of the same color, say gray, and therefore to find the right folder she will have to read the tags until she finds the one that says “Dynamic Semantics” and then give me the folder. In this case, information about color is not relevant for the task that my colleague has. Let us now imagine that, in the exact same dialogical situation, my folders are colorful, and that my colleague knows my “Dynamic Semantics” folder is the red one; browsing through my shelves in my office, she’ll look for the red folder; color information will be in this case salient for the task at hand. This has a lot to do with the fact that the color of an object can be of some relevance or not depending on the situation. When browsing the room looking for an object, different characteristics can be relevant and therefore acquire salience.

There’s no intention here to directly compare a perceptual task like that described in the study of Winawer and colleagues to the described situation; the two tasks clearly involve different levels of explicitness and entail different relationships between the attribute color involved and the rest of the representation; however, the point is to embrace the intuitive idea that information about certain features of a determined object can be more or less salient and relevant depending on the task at hand. What these classical examples in pragmatics show is that, in communication, features associated with an object can acquire relevance and salience depending on the situation at hand. In these communicative situations, arguably, mental representations are employed to “solve” the comprehension or production task. In the case of the red folder, different attributes acquire relevance.

---

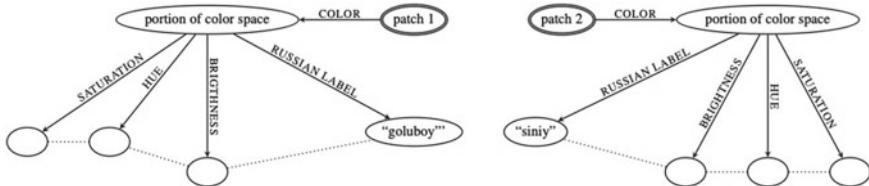
<sup>9</sup>Thanks to the anonymous reviewer for bringing my attention to this fact.

This kind of idea is not only intuitively plausible, but also what underlies research enterprises in psycholinguistics that are meant to assess what the relationship between concepts and their components is; for instance, studies like Redmann and colleagues (2014) investigate the activation of color attributes in high color-diagnostic concepts (like, for instance, bananas). Studies like this focus on language production; however, the idea is that concepts can be treated as complex structures whose different components can be “activated” depending on the situation. Moreover, it is assumed that definite relations among attributes and nodes in a frame exist, the idea being that the activation of a conceptual component can potentially facilitate the activation of other parts of the concept.

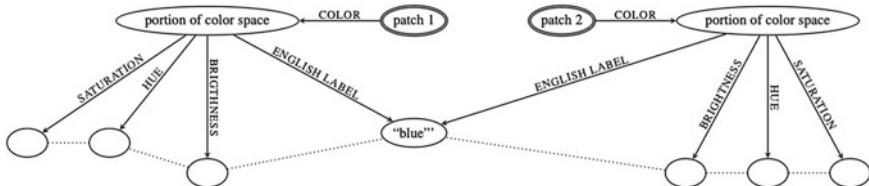
Another analogy will help clarify the position. Consider my own representation of DOG. Presumably, it entails different kinds of attributes encoding several kinds of information - purely perceptual, verbal, and so on. Approximately, a frame representation of DOG for me might include not only information about basic dog attributes - such as for instance number of legs, fur, eating habits, and so on, but also plenty of information about Nala, my dog, about other dog encounters that I had in the past, about my grandma’s dog that I got to know when I was very young, about the names for dogs I’ve heard most often when in Italy, and so on. This entire repertoire of information, however, does not need to be recruited every time I have to activate my dog representation in a communicative situation; it’s reasonable to think, on the contrary, that this only happens when certain kind of information is required, or relevant, for a given task - namely, the one I am performing, whatever this might be. Depending for instance on the communicative situation, I will need to recruit different kinds of knowledge.

Let us now apply this understanding of concepts and attributes within them to the main focus of the paper, trying to put the pieces together. The debate is open as far as how lexical information enters the conceptual domain, as described above; the question of how linguistic representations and non-linguistic ones interact is precisely the kind of question that, after all, guides the debate about Linguistic Relativity. On the other hand, if one assumes that information about how certain perceptual features can be linguistically coded in different ways (hence, that we can assume the presence of attributes-like structures like the LABEL one and that the value can change) and that conceptual components can be recruited *according to the situation and the context at hand*, it is natural to assume that the linguistic information can or cannot be activated and recruited, depending on the context. The modalities and circumstances of this activation, then, would need to be investigated.

A case like that of Winawer seems to suggest that conceptual representations of colors, and consequently their labels, can be used and activated during a perceptual task; one of the possible interpretations of the results is that, while English speakers operate comparing different perceptual inputs without activating linguistically coded representations, Russian speakers use a different strategy, namely they employ color concepts and their labels; at least that’s what seems to be suggested by the difference in performance. Crucially, however, this kind of strategy seems to be replaced by the same strategy English speakers employ, in case of linguistic interference: somehow, then, performing another linguistic task “blocks” or inhibits the label-influenced



**Fig. 4** Winawer’s task in frames: Russian



**Fig. 5** Winawer’s task in frames: English

strategy. Given the fact that the task is still possible for English speakers, this is clearly not something that prevents them from performing the task, regardless of the presence of color labels. What this study seems to suggest, then, is that recruiting or not recruiting linguistic information can depend on the type of task: in this sense, the choice of strategy is flexible.

Let us try and represent this in frames again with Figs. 4 and 5.

A plausible explanation that is easily representable in frames is that the task is solved by the Russian speakers by comparing two different nodes including linguistic information. This strategy is not available in the case of English speakers, since there is only one node containing linguistic information available; therefore, a strategy based on comparing, for instance, visual patterns in SATURATION, HUE and BRIGHTNESS is used. Russian speakers can then shift to the same strategy when the label attribute is unavailable- i.e. in within-category trials.

To reiterate: this means assuming that it is possible to draw a parallelism between concepts like BANANA and concepts like BLUE; in other words, assuming that it makes sense to consider an attribute like *label* (in language x) to be something that pertains to the representation of both. In a sense, this is the first tenet of the model presented here. The second tenet is that a mental representation can be considered as a structured file where not every part gets activated every time the concept is evoked; instead, the amount and the kind of information that will be used in the task at hand will vary according to task constraints, context and possibly other factors. Finally, a point that has been stressed while presenting the view is that different kinds of information, of perceptual and not perceptual nature, can be incorporated in the same mental representation.<sup>10</sup>

<sup>10</sup>This is clearly not the only available theory. An alternative account can for instance be found in Newen (2011) A thorough comparison between the two views would be fruitful but would go beyond

Arguably, more research has to be done in this direction, as the issues are multiple and complex. However, it should be clear that results of studies like that of Winawer or Roberson should be considered as interesting because they fit into an account of cognitive processes manipulating representations in a flexible, task dependent way, where different information is recruited according to what is useful for the task at hand. In Winawer's case, paradigmatically, linguistic labels seem to play the role of facilitators for the task at hand, or at least to make a difference when recruited. Phrased using the vocabulary introduced until now, this implies assuming that there are complex interactions among linguistic information and perceptual information which are functionally connected and can be differently employed. Frames are just one way to represent this kind of relation: however, they help in seeing how data such as that presented, more than settling the debate about language relativism, should suggest to see it in another light. A difference between "shallow" and "deep" Whorfianism ceases to be relevant, once one assumes that the kind of information that has to be considered when modeling mental representation can be of different kinds (linguistic and perceptual, for instance) and that this kind of information interacts in complex ways: the fact that effects of language categorization on cognitive tasks vary depending on context and task demands seems to point towards an understanding of mental representations precisely in this direction.

So far, it has been argued that a view of mental representations that involves flexible use depending on the task at hand can be represented efficiently in frames and that it has a good chance to be related to a model of how representations are used in communication. However, a few steps are still needed. In the Russian-English speakers example, what we apparently have is the use of two different strategies for performing the task: however, there is still no direct evidence in favor of considering "LABEL" as an attribute that gets activated depending on the task. For all we know, the strategy employed by English speakers (and by Russian speakers when linguistic interference is present) might not include any kind of conceptual activation. Participants might be comparing perceptual input, solving the task on the basis of this comparison, and using a strategy based on labeled mental representations instead when two different color terms are present: this suggests switching between strategies, but does not support necessarily the idea that the linguistic information in a concept can be activated or not depending on the situation. I think this is a viable option, as will be argued below. In order to push further Lalumera's suggestion, to consider the compatibility of the color terms evidence with a more dynamic picture

---

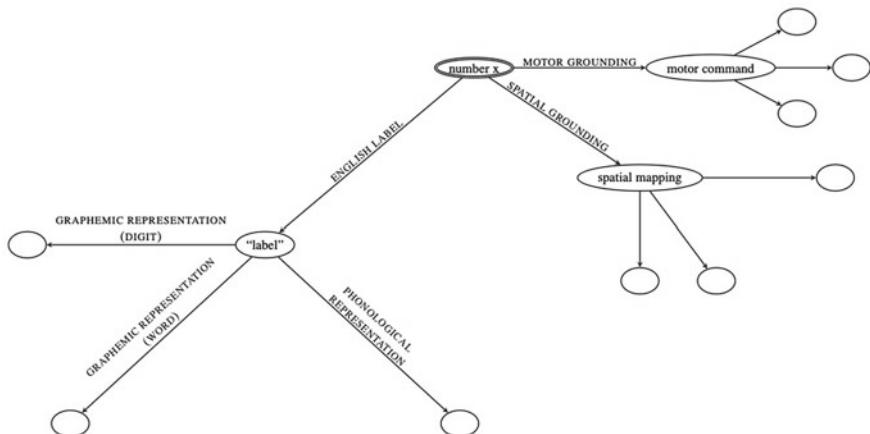
the aim of the paper. Two basic differences are however to be noticed; firstly, Newen adopts a model where relations between conceptual parts are not spelled out in terms of functional relations like in frames. Secondly, he makes a distinction between two different concepts: RED referring to the property of being red and RED EXPERIENCE referred to the property of having a red experience, where the information contained in the first can be integrated in the latter, albeit not as a defining component. I believe this idea could be integrated in a frame network, but this would require further investigation.

of mental representations, it is necessary to go a few steps further. To get there, we will consider now a different example from another conceptual domain before turning to colors again.

## 5 A Brief Excursus into Another Conceptual Domain: Counting and Motor Representations

As argued so far, in the case of cross-linguistic evidence for color terms, the debate has focused a lot on whether effects are to be considered “just” shallow and temporary or “deeper”. In the context of embodied cognition, something very similar has happened, in a somehow opposite direction. Embodied semantics is concerned with the role of motor and perceptual representations in conceptual units, the idea being that is worth exploring the multimodality of mental representations or, in other words, the role that sensory modalities play in their structure, use and retrieval. One of the battle grounds in the embodied cognition debate has always been that of abstract concepts: even if it’s more or less accepted that motor and perceptual information can have some relevance as long as concrete concepts are concerned, the same does not hold for concepts that, intuitively, have less to share with perception, hence abstract concepts. Moreover, one common argument against embodied cognition lies in the idea that, even when perceptual and motor resources are recruited during semantic processing, this is only a somehow shallow “cascade effect” that has nothing to do with “deeper” conceptual processing (Mahon and Caramazza 2008).

In the context of research regarding representations of numbers, which are considered quite abstract, there have been several attempts to connect numbers and counting to the more (supposedly) concrete domain of space, the idea being that abstract concepts like mathematical ones are mapped to more concrete representations like spatial ones, which is what guarantees their being “grounded” in experience. In a famous study run by Dehaene and colleagues (2019), the so called SNARC (Spatial Numerical Association of Response Codes) effect was described: large numbers elicited rightward response and small numbers leftward ones, meaning that small numbers were classified faster with the left hand and bigger digits were classified faster with the right hand. Since similar effects were found as long as the vertical axis is concerned (up for bigger digits and down for smaller ones), this kind of idea was investigated in a number of other studies. A particularly interesting one is that by Pecher and Boot (2019). The task was to judge the magnitude of numbers in comparison with other digits: the stimulus was a digit that was located congruently or incongruently with the image schematic location of the number (left for smaller digits, right for bigger ones). In the concrete contexts, participants had to say whether the digit was bigger or smaller than the one in concrete sentences (“The man read two books a day”). In the abstract context condition, the digits were to be compared to other numbers. The idea was to test whether the congruent spatial condition was facilitating the task or not, which ended up being true *only for the concrete context*.



**Fig. 6** Frame for a number

Regardless of the debate about embodied cognition, which is vast and complex, the result is interesting because it has been used to argue against the idea that spatial representations are relevant for number processing *because they only appear to be used in certain processing contexts*. This is somehow very similar to what happens in the color labeling debate: even here, the key of the arguments lies in the fact that certain kind of information is only thought to be relevant in determined contexts and tasks. However, this is hardly enough to say that the positive result (the facilitation effect in the concrete condition) is not interesting: on the contrary, it suggests that different processes are going on linking different kinds of information depending on the task at hand. Moreover, the result goes hand in hand with theories of embodied cognition like that proposed by Barsalou (2008), where the role of motor and perceptual representations and that of linguistic ones *varies depending on the type of task*, but where both have a crucial role in conceptual representations.

Let us look at a possible frame for a concept of a number in Fig. 6.

Different kinds of attributes are present, comprising different kinds of information. A number has a label, which implies a phonological representation *and* a graphemic one and, in this picture, includes spatial mapping information and possibly motor grounding (lots of the research regarding grounding of number has focused on finger counting).

A frame like that in Fig. 6 *does not imply* that motor grounding and spatial information are always recruited when the concept of a number is evoked. On the contrary, it is conveniently compatible with the view of mental representations that has been presented so far and with the idea that different attributes can be recruited depending on the situation at hand. Let's consider the experiment reported: in one condition (the concrete one), spatial information seems to be relevant, since the subjects' performance changed depending on whether the spatial information was congruent with the magnitude of the numbers or not. One can then assume that the attribute named

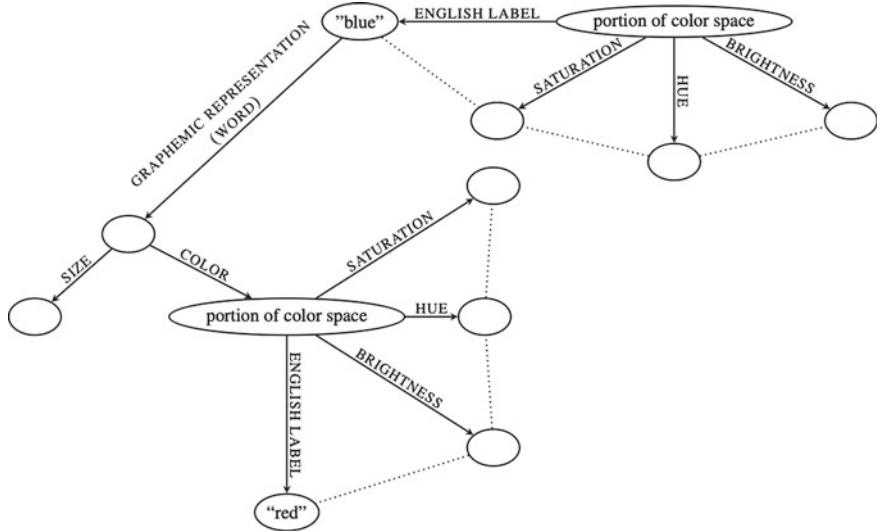
here “spatial grounding” was then evoked and recruited. The same does clearly not apply to the abstract condition: in this case, the spatial information did not seem to be relevant, since the performance did not change depending on the congruency of the position. This, more than speaking for an alleged scarce relevance of the spatial mapping, seems to suggest that some other kind of information was relevant for the task: for instance, the graphemic representation was probably employed. Lacking a concrete context for the digits, the task was performed using a different strategy, which probably included in this case comparing the graphemic representations of the numbers: this is another kind of information, namely visual. Even in this case, there is a switching of strategies. However, this time, it is plausible to think that different parts of the involved mental representations are recruited. Depending on task demands and conditions, different parts of the representations are relevant, and different attributes are activated. The frame captures the multi-modal nature of the concept and the flexibility that underlies its use.

## 6 Back on Colors: Stroop Task And Language-Perception Interface

Let us then come back to colors now, and consider another set of evidence that is often discussed, namely the Stroop effect. The phenomena was investigated for the first time in 1935 (Stroop 1935), and very often recreated. In the traditional set up, color words are printed in either congruent or incongruent ink (e.g. the word blue is printed either in blue or red, for instance), and participants are instructed to name the color of the ink used for printing and to ignore the meaning of the word. Typically, the task is quite difficult and the incongruent trials cause a significant delay in reaction times.

Let us think about a possible frame (Fig. 7) describing the situation in the same terms that have been spelled out above:

Even in this case, there is a graphemic representation of the *English label* that can be included in the mental representation. Being a graphemic representation, it is perceived by the viewer; hence, it makes sense to include perceivable attributes in the frame. The font will have a size and a color, for instance; only the latter is then relevant for the task at hand, which is the individuation of the color. The label that is represented on paper, however, also has a clear connection with a color concept, that includes a portion of color space (and therefore has determined attributes). Now, what can happen in such a representation is that the two portions of color space involved have different values in terms of *saturation*, *brightness* and *hue* i.e. that they identify a different color, possibly named differently. The mental representation becomes, in this sense, more complex and can therefore be the reason why processing costs actually become higher: having to produce a response based on the label given to a color concept, and being the case that two different labels and two different concepts are evoked and involved, the task becomes difficult to solve. Note that the participant



**Fig. 7** Frame for a Stroop task (incongruent colors)

does not perceive the label “red” anywhere; however, an attribute is evoked and activated and the task gains complexity and potentially makes it easier to produce mistakes. Having two nodes of the same kind, with the same sort of information, makes it harder to process it, since there is conflicting information regarding the label involved in the task. In a way, this is the opposite of what happens in the case of the blue houses; since the task is not a discrimination one, but rather one where one label has to be produced, the presence of two different nodes of the same kind delays solving the task.

## 7 Conclusions and Open Questions

In the present paper, a way to model color representations has been proposed that represents them as complex structures used in perception tasks and communicative tasks in a flexible way. The view, as stressed above, is not meant to disprove or support Whorfian-like hypotheses. Rather, the model shows how task requirements shape conceptual retrieval, and how complex representations can be used flexibly in the context of specific tasks in a way that is compatible with the evidence regarding color terms and perceptual tasks presented. Lalumera’s suggestion, to consider the idea that “shallow” effects of language labels on non linguistic tasks are still interesting if one does *not* assume mental representations to be rigid units, is here accepted and pushed a bit further: it has been argued that what the evidence suggests is, as a matter of fact, that a view of mental representations that integrates several kinds of information,

recruited flexibly and task-dependently, is indeed able to potentially account for the findings. This idea is implemented in terms of functional attributes representing linguistic information. This is embedded in a view where mental representations are modeled in terms of different kinds of information as functionally integrated in a complex structure, which is what results like that of Pecher and Boot actively seems to suggest and what can be potentially modeled in the Stroop task case.

The presented evidence clearly only gives some clues about how determined mental processes are affected by linguistic labels for perceptual information and about how this can be modeled. The limited set of examples, moreover, can only partially be considered decisive, and the advanced proposal has to be integrated in a full blown theory of frames. The ultimate goal of such a proposal, moreover, would be to have a empirical paradigm that addresses the specific hypothesis regarding the structures of the representations involved. However, the fact that the model seems to be potentially able to accommodate evidence from different research fields is encouraging as far as the possibility to have a better understanding of how perceptual and linguistic information interaction in complex mental representations goes

**Acknowledgements** First, I would like to thank the organizers of the ‘Cognitive Structures: Linguistic, Philosophical and Psychological Perspectives’ Conference and the editors of the current volume. My special thanks go to my supervisor, Prof. Gottfried Vosgerau, for the support and help with the research, and to the whole SFB 991 team for the fruitful input and discussion and for the always stimulating environment. Thanks in particular to Kurt Erbach for revising the manuscript and providing constant feedback, Natalja Beckmann and Alexandra Redmann for discussing some versions of the frames with me, and to Katja Gabrovska for the templates. This research was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft) CRC 991, Project D02.

## References

- Barsalou, L.W. (1992). Frames, Concepts and Conceptual Fields. In: A. Lehrer, E.F. Kittay (Eds.), *Frames, fields and contrasts*. Hove and London: Erlbaum, Hillsdale, 2174.
- Barsalou, L.W., Santos, A., et al. (2008). Language and simulation in conceptual processing. *Symbols and Embodiment: Debates on meaning and cognition*, Oxford University Press. Chap. 13.
- Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. University of California Press.
- Carruthers, P. (1996). *Language*. Thought and Consciousness: Cambridge University Press.
- Casasanto, D., Lupyan, G. (2019). All concepts are ad hoc concepts. In E. Margolis, & Laurence S. (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 543–566). Cambridge: MIT Press.
- Dehaene, S., Bossini, S., & Giroux, P. (2019). The mental representation of parity and number magnitude. *Journal of Experimental Psychology*, 122.
- Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. (2007). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and Language*.
- Lalumera, E. (2014). Whorfian effects in color perception: Deep or shallow? *Baltic International Yearbook of Cognition, Logic and Communication*, 9, 1.
- Löbner, S. (2015). Functional concepts and frames. In T. Gamerschlag, D. Gerland, R. Osswald, & W. Petersen (Eds.), *Meaning, frames, and conceptual representation. Studies in language and cognition 2* (pp. 13–42). Düsseldorf University Press.

- Macpherson, F. (2012). Cognitive penetration of colour experience: Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research*, LXXXIV(1).
- Mahon, Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology*, 102, 59–70.
- Newen, A. (2011). Phenomenal concepts and mental files: Phenomenal concepts are theory-based. *Philosophia naturalis*, 47(48), 155–183.
- Newen, A., & Vetter, P. (2017). Why cognitive penetration of our perceptual experience is still the most plausible account. *Consciousness and Cognition*, 26–37.
- Pecher, D., & Boot, I. (2019). Numbers in space: Differences between concrete and abstract situations. *Frontiers in Psychology*, 2.
- Rayo, A. (2011). A plea for semantic localism. *Nous*, 47(4), 647–679.
- Redmann, A., FitzPatrick, I., Hellwig, F., & Indefrey, P. (2014). The use of conceptual components in language production: An ERP study. *Frontiers in Psychology*, 5, 363.
- Roberson, D., Pak, H., & Hanley, R. (2008). Categorical Perception of color in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition*, 107, 752–762.
- Sapir, E. (1983). In *Selected Writings of Edward Sapir in Language, Culture, and Personality*. University of California Press.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2006). Russian blues reveal effects of language on color discrimination. *PNAS*, 104, 7780–7785.
- Wolff, P., & Holmes, K. (2010). Linguistic relativity. In *Cognitive Science* (Vol. 2, pp. 253–265). Cambridge: Cambridge University Press.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Implicatures and Naturalness



Igor Douven

**Abstract** Pragmatics postulates a rich typology of implicatures to explain how true assertions can nevertheless be misleading. This typology has been mainly defended on the basis of a priori considerations. We consider the question of whether the typology corresponds to an independent reality, specifically whether the various types of implicatures constitute natural concepts. To answer this question, we rely on the conceptual spaces framework, which represents concepts geometrically, and which provides a formally precise criterion for naturalness. Using data from a previous study, a space for the representation of implicatures is constructed. Examination of the properties of various types of implicatures as represented in that space then gives some reason to believe that most or even all types of implicatures do correspond to natural concepts.

**Keywords** Conceptual spaces · Implicatures · Multi-dimensional scaling · Natural concepts · Pragmatics

Linguists and other language researchers customarily distinguish between syntax, semantics, and pragmatics, where (roughly) the first pertains to the ways words can and cannot be combined into sentences, the second to word and sentence meaning, and the third to language use. This paper is concerned with a question central to pragmatics, specifically with the scientific status of so-called implicatures, which play a key explanatory role in this field. More specific still, we are interested in the question of whether all types of implicatures that the current literature distinguishes between are *natural* concepts, where the notion of a natural concept will be understood as defined by researchers working on psychological spaces. The question is important insofar as only natural concepts deserve a place in mature scientific theories (Lewis 1983; Boyd 1991).

To address this question, we use data from a study reported elsewhere (Douven and Krzyżanowska 2019) to construct a psychological space for the representation of implicatures. In that space, we examine the properties of various types of impli-

---

I. Douven (✉)  
SND/CNRS/Sorbonne University, Paris, France  
e-mail: [igor.douven@paris-sorbonne.fr](mailto:igor.douven@paris-sorbonne.fr)

catures, with a special interest in seeing whether they satisfy an important criterion for naturalness (convexity—see below) as proposed in the psychological spaces literature. The outcome will be seen to provide some support for holding that most or even all types of implicatures do correspond to natural concepts.

## 1 Theoretical Background

The basic insight at the root of pragmatics is that we can mislead our audience not only by telling lies, but also by telling nothing but the truth. Suppose someone asserts,

- (1) President Obama has one daughter.

The assertion is true yet misleading, given that it suggests that Obama has *exactly* one daughter—which is false. What is suggested is not asserted, but it is nonetheless conveyed due to a normally warranted presumption of a kind of cooperativeness that goes beyond merely telling the truth. In the present example, we may suppose that the speaker was in a position to assert, and could with just as much effort have asserted, that Obama has *two* daughters, which would have been true as well but would in addition have been more informative. Precisely because we expect each other to be cooperative in this kind of way—to try to make our contributions to a conversation not only true but also relevant, clear, and informative—a person unaware of how many daughters Obama has would be justified to infer from an assertion of (1) that he has exactly one daughter. That Obama has *exactly* one daughter is said to be an *implicature* of (1), whose *semantic* content is only that Obama has one daughter, possible among many more.

There exist a number of different typologies of implicatures, which are partly independent of each other. One broad division is that between *conventional* and *conversational* implicatures, where the former are said to arise due to the meaning of specific words, and the latter due to the context in which an assertion is made. For instance, the word “although” in

- (2) Although Obama won a second term as president, dolphins are mammals.

Suggests the existence of a contrast between the two conjuncts in this sentence (which strikes us as wrong, given that the conjuncts appear unrelated). On the other hand, there is no single word in (1) that might lead a hearer to think that Obama has *exactly* one daughter. *That* suggestion can arise for the reason mentioned above: because we would normally assume that (1) is the strongest statement the speaker can make regarding the number of daughters Obama has. Indeed, there are conversational contexts where this assumption would not be warranted. For instance, if it has just been asserted that anyone who has at least one daughter qualifies for a certain special government program, we would not interpret an assertion of (1) as suggesting that Obama has exactly one daughter. Rather, we would take the speaker’s point to be that Obama meets the requirement for the government program.

This brings us to second distinction. We just said that although an assertion of (1) would, in normal circumstances, implicate that Obama has exactly one daughter, there are circumstances in which this implicature would not arise. Grice (1989, p. 37f) calls implicatures of this type “generalized conversational implicatures.” He differentiates them from what he calls “particularized conversational implicatures,” which arise only in specific conversational contexts. For instance, if we are at a party and you ask me what time it is, you may interpret my assertion of

- (3) The guests are leaving.

As indicating that it is already late, even if asserting (1) normally does not engender this suggestion.

It is fair to say, though, that most attention in the literature has gone to a sub-typology of conversational implicatures which is based on the various types of expectations—each brought about by the overarching expectation of cooperativeness—that the implicatures exploit. For instance, the aforementioned implicature of (1) is said to be of a scalar type, because we can represent numbers (e.g., numbers of children) on a scale, and the expectation of informativeness then requires that we go as far out on that scale as is warranted by our evidence. So someone’s asserting (1) implicates that she knows, or has good evidence for believing, that Obama has exactly one daughter. By contrast, someone asserting that

- (4) Kate Middleton gave birth to a son and she married Prince William.

Is offending the expectation that we report events in an orderly fashion, which in this instance means: in the order in which they occurred. Thus, the obviously wrong implicature generated by an assertion of (1)—that the event mentioned first also happened first—is said to be of an order type.

Scalar implicatures have given rise to a further sub-typology, this one being based on the different scales that can underly the production of these implicatures. The main subtypes are the *quantificational implicatures*, which involve a scale of quantifiers (e.g., some–many–most–all); the *gradable adjective implicatures*, which exploit some scale of adjectives that can apply to differing degrees (e.g., soft–audible–loud–blaring); the *ranked ordering implicatures*, which involve orderings (like beginner–intermediate–advanced); and the *cardinal number implicatures*, which involve some cardinal number scale, as in our example (1).

This paper will focus on the typology which starts by branching off the conversational and conventional implicatures and which then has the further branches for the conversational implicatures described in the previous two paragraphs. This typology has been mainly defended on the basis of a priori considerations, more specifically on what are sometimes called “linguistic intuitions.” However, such intuitions are known to be not always reliable. Indeed, while the said typology is still part of mainstream pragmatics, parts of it have been contested. For instance, some authors deny that sentences like (1) carry the “exactly  $n$ ” reading as a matter of implicature, claiming that, rather, the “exactly” reading is part of the semantics of numerals (see Scharten 1997 and Breheny 2008). And Bach 1999 has argued that the belief in the existence of conventional implicature rests upon a myth.

Bach's arguments have in turn been challenged (e.g., Potts 2005) and in any event my aim is not to question the reality of any part of the aforementioned typology. Rather, I am interested in the metaphysical status of the various types that occur in it. It has often been said that we do not just want scientific theories to be predictively accurate, but also want them to inform us about what, deep down, underlies the phenomena (e.g., Psillos 1999). And that requirement can be satisfied only if these theories "carve nature at its joints," that is, only if their core concepts are *natural* ones (Lewis 1983). Against this background, the question I am asking is whether the above typology latches on to some independent, fundamental reality. Do, for instance, so-called order implicatures constitute a *natural* class of implicatures? More generally, are *all* types of implicatures natural? Or better perhaps, if Lewis (1983) is right that naturalness permits of degree, are they all *equally* natural?

To address these questions, we need some understanding of what it takes for a concept to count as natural. It has been argued that a concept is natural if it figures in one or more laws of nature (e.g., Putnam 1983). But this is problematic, given that it is hard to say what makes a regularity a law of nature (or otherwise) without making reference to natural concepts (Douven and van Brakel 1998). To characterize naturalness of concepts, it is actually more helpful to turn to recent work on conceptual spaces, in which a criterion for distinguishing natural from nonnatural concepts has been proposed that is backed by a considerable amount of experimental evidence.

We will construct a conceptual space later on, and will then go into details. For now, it suffices to say that a conceptual space is a one- or multidimensional metric space, where the dimensions represent fundamental qualities that items can have to varying degrees and with respect to which they can be compared to each other. Distances in such spaces are supposed to be inversely related to similarities: the greater the distance between (the representations of) two items in a given space, the more dissimilar the items are in the respect represented by the space. For example, CIELAB space is a three-dimensional Euclidean color space, and distances in the space are meant—and have been shown—to predict accurately how similar people will judge different shades to be: the closer two shades are in CIELAB space, the more similar they tend to appear to human observers (Fairchild 2013). Many other conceptual spaces are known in the literature, and although the best-known ones all pertain to perceptual concepts (next to color spaces, such as CIELAB, there are vowel spaces, odor spaces, taste spaces, etc.), more recently conceptual spaces have been developed for more abstract concepts, including moral, epistemic, and scientific concepts.

What makes conceptual spaces especially valuable is that they allow us to represent concepts geometrically, as regions in some given space. Thereby, the study of concepts becomes both formally rigorous and empirically testable. For instance, the concept of redness can be thought of as a region in CIELAB space, which means we can carry out all sorts of mathematical operations on it—like measuring its volume—and at the same time use it for conducting all sorts of experimental work (e.g., concerning the nature of vagueness: see Douven et al. 2013; Decock and Douven 2014; Douven and Decock 2017; Douven et al. 2017; Douven 2018).

If concepts are regions in conceptual spaces, is any region in any conceptual space a concept? “Concept” is, to a high degree, a term of art, and so we are free to answer this question in the positive. However, the more worthwhile question is whether any region represents or could represent a *natural* concept. And it takes little imagination to appreciate that now the answer is definitely negative. In color space, there are infinitely many regions that contain all the colors in the rainbow. Surely such regions represent gerrymandered rather than natural concepts.

Now that we can think of concepts formally, can we also distinguish formally between those regions that represent or can represent natural concepts and those that can *not*? Gärdenfors (2000, p. 71) proposes a topological criterion, which he calls

Criterion P: A *natural concept* is a convex region of a conceptual space,

where a region  $\mathcal{R}$  is convex if and only if, for any pair of points  $x, y \in \mathcal{R}$ , if  $z \in \overline{xy}$  then  $z \in \mathcal{R}$ . As Gärdenfors (2000, p. 70) explains, Criterion P can be thought of as a principle of cognitive economy, given that “handling convex sets puts less strain on learning, on your memory, and on your processing capacities than working with arbitrarily shaped regions.” He also cites important empirical work on color naming which shows that color concepts like BLUE, RED, GREEN, and so on, which we tend to regard as *natural* color concepts, all form convex regions in CIELAB space (see also Jraissati and Douven 2018). Douven (2016a) presents further empirical evidence for Criterion P, showing that the concepts BOWL and VASE come out as convex in the appropriate shape space.

Whereas Criterion P is a plausible *necessary* condition for natural concepts, it is debatable whether it is also *sufficient*.<sup>1</sup> Gärdenfors (2000, p. 70) already expressed doubts on this point, and Douven and Gärdenfors (2018) argue explicitly that further conditions are needed to single out the natural concepts. However, in addressing the question of whether all types of implicatures are equally natural concepts, we will content ourselves with considering whether the various types of implicatures, when represented in a conceptual space we are about the construct, satisfy Criterion P. If some fail to do so, that is an indication that they are not natural concepts. And if some or all do satisfy the criterion, that is at least some evidence for holding that they *are* natural concepts.

To build the requisite conceptual space for representing types of implicatures, we need input data. The data we are going to use are taken from a study reported in Douven and Krzyżanowska (2019). We briefly describe the data in the next section, and then go on to construct a conceptual space in Sect. 3.

---

<sup>1</sup>There has been some discussion about whether Criterion P is even necessary. See Gärdenfors (2018) and references given there.

## 2 Input Data

Douven and Krzyżanowska (2019) were interested in three questions, all related to the semantics–pragmatics interface. First, they sought to investigate empirically whether ordinary speakers’ responses to true but supposedly pragmatically infelicitous sentences—true sentences that generate a false implicature—are in line with linguists’ and philosophers’ ideas about how semantic and pragmatic aspects of language are to be sorted. Specifically, they were interested in whether people reliably distinguish between the *truth* and the *assertability* of sentences in a way that accords with mainstream thinking in linguistics and philosophy.

Second, Douven and Krzyżanowska (2019) were interested in possible differences in responses brought about by the various types of implicatures. For instance, might people systematically deem true sentences generating false *conventional* implicatures *more* unassertable than true sentences generating false *conversational* implicatures? Might the different types of conversational implicatures be evaluated differently in this respect?

And third, they were interested in individual differences among participants. Previous research (Spychalska, Kontinen, and Werning 2016) had suggested that some people are more inclined to judge the truth values of sentences purely on the basis of what according to theorists are the semantic contents of those sentences, whereas other people might base their truth judgments also, at least to some extent, on the sentences’ pragmatic aspects, so that they might be more inclined to judge a true sentence with a false implicature as *false*.

To investigate these questions, Douven and Krzyżanowska used materials consisting of the 24 items listed in Table 1 together with a great variety of filler items which were meant to conceal from the participants the purpose of the study. The test items were meant to generate six types of false implicatures, where each type was instantiated by four different sentences: quantificational implicatures (items 1–4); gradable adjective implicatures (items 5–8); ranked ordering implicatures (items 9–12); cardinal number implicatures (items 13–16); temporal order implicatures (items 17–20); and conventional implicatures (items 21–24).

In both studies reported in Douven and Krzyżanowska (2019), the participants were divided into three groups, where participants in one group were asked about the items’ *truth*, participants in a second group were asked about the items’ *assertability*, and participants in the remaining group were asked about the items’ *believability* (the questions about believability were related to a secondary research goal, which we leave aside here; see Douven 2010, 2016b, and Douven and Krzyżanowska 2019). The difference between the two studies was that participants in the first were always asked to give yes/no answers, whereas participants in the second study were asked to indicate on a 7-point Likert scale the extent to which they agreed that an item was true/assertable/believable.

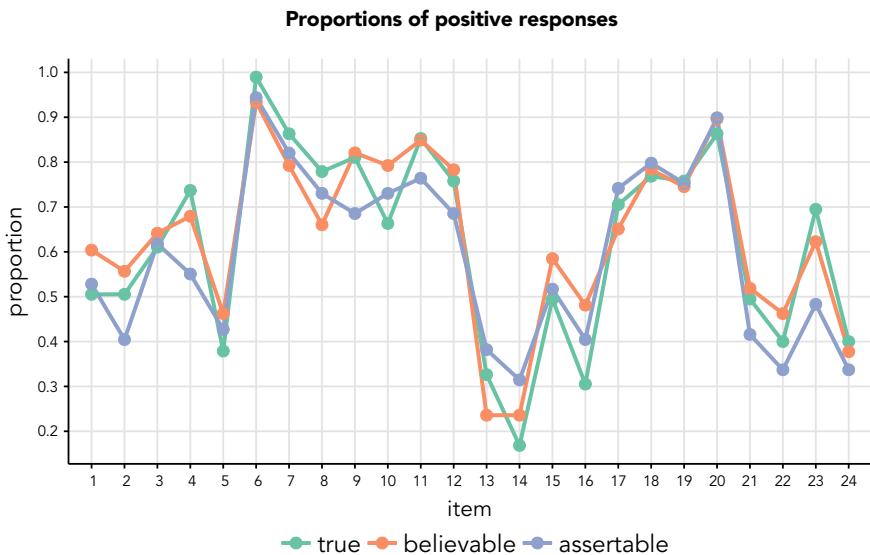
As for the first research question, neither study revealed any significant differences among the responses from the three groups (nor were there significant differences between the two studies). Figure 1 presents the proportions of positive responses from

**Table 1** Items used in the studies reported in Douven and Krzyżanowska (2019)

1.	Some patches are blue. <sup>a</sup>
2.	Some roses are flowers
3.	Most patches are red. <sup>b</sup>
4.	Most laptops are computers
5.	The tiger finds the boy's cereal moderately sweet. <sup>c</sup>
6.	The female basketball player Margo Dydek (7 ft 2 in./2.18 m) was tall for a woman
7.	Bill Gates is relatively rich
8.	On the North Pole, winter temperatures are somewhat cold
9.	In the UK, people over the age of 85 have the right to retire
10.	In principle, all American citizens over the age of 25 have the right to vote in federal election
11.	In the UK and the US, children under the age of 15 are prohibited from buying hard drugs
12.	In the US, people who earn more than \$200,000 a year are obliged to pay taxes
13.	Alfred Hitchcock made two movies
14.	President Obama has one daughter
15.	In the last Olympic games, the US won four medals
16.	At the height of its power, Great Britain owned 12 ships
17.	The tiger looks for the bread in the toaster and the boy puts a piece of bread into the toaster. <sup>d</sup>
18.	Princess Diana died in a car accident and she divorced Prince Charles
19.	The man comes up with a bogus answer and the boy asks how the load limit on bridges is determined. <sup>e</sup>
20.	Kate Middleton gave birth to a son and she married Prince William
21.	Although Prince William had fallen in love with Kate Middleton, the 2014 Winter Olympics will be in Russia
22.	<i>Harry Potter and the Sorcerer's Stone</i> was a box office hit, therefore Obama is the president of the US
23.	Although Obama won a second term as president, dolphins are mammals
24.	Mitt Romney lost the 2012 presidential election, therefore U2 is a rock band

<sup>a</sup>Shown with a series of only blue patches. <sup>b</sup>Shown with a series of only red patches. <sup>c</sup>Shown with a comic strip in which a tiger is seen finding a boy's cereal extremely sweet. <sup>d</sup>Shown with a comic strip in which a boy first puts bread in a toaster and then a tiger looks into the toaster. <sup>e</sup>Shown with a comic strip in which a boy first asks the question and then the man answers it

the first study, which shows how close the responses from the three groups were to each other. The graphs of the mean responses from the second study, not shown here, are virtually indistinguishable from those shown here; see Douven and Krzyżanowska (2019). So, as far as these results go, it hardly appears to matter whether we ask people to judge the truth, believability, or assertability of a sentence that is true according to standard semantics but that generates a false implicature. More generally, Douven and Krzyżanowska (2019) found no evidence that the semantics–pragmatics divide,

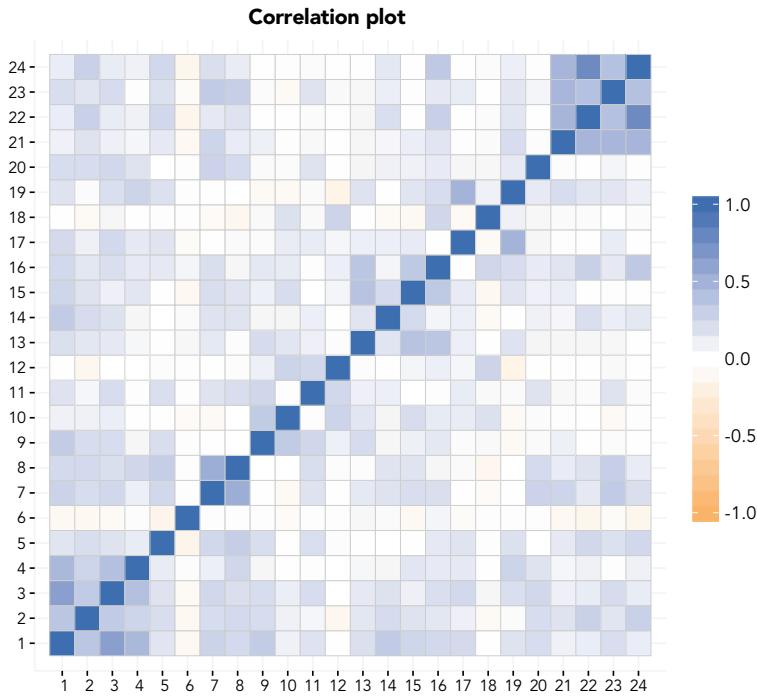


**Fig. 1** Proportions of positive responses per item from the first study in Douven and Krzyżanowska (2019); labels refer to the numbering of items in Table 1

however useful from a theoretical perspective perhaps, is reflected in how ordinary speakers tend to evaluate sentences like those in Table 1.

As stated above, Douven and Krzyżanowska (2019) were also interested in possible differences in responses due to the various types of implicatures generated by their materials. Just eye-balling the results in Fig. 1, it appears that proportions of positive responses tend to be in the same range for each type separately, but not so much across types. In line with this, Douven and Krzyżanowska's analysis revealed a significant effect of type of implicature on the responses. They again obtained the same result for the responses from their second study. Hence, the answer to their second question was positive.

For the third question—whether participants can be split into logical responders and pragmatic responders—they looked at the correlations between the responses for any pair of items. If a division between logical and pragmatic responders exists, then at a minimum one would expect these correlations to be rather high: some participants—the supposedly logical responders—would then tend to judge all items in Table 1 to be true, while others—the supposedly pragmatic responders—would tend to judge all those items to be false. But that turned out not to be the case. Figure 2 is reproduced from Douven and Krzyżanowska (2019) and shows the correlations among the “truth” responses from the first study; the correlations from the second study were essentially the same. It is clearly visible that, whereas both the responses to the quantificational items and the responses to the conventional items correlate amongst themselves, they do not even moderately correlate with most of the other



**Fig. 2** Correlations among “truth” responses from the first study in Douven and Krzyżanowska (2019); labels refer to the numbering of items in Table 1

items, nor do the responses to those other items tend to correlate even moderately among themselves.

Given that in no interesting respect were there significant differences between the two studies reported by Douven and Krzyżanowska, we in the following consider only the data from the first study.

### 3 Building an Implicature Space

In Sect. 1, we mentioned that, whereas most conceptual spaces to be found in the literature are for *perceptual* concepts, there is nothing that prevents us from constructing spaces for other types of concepts, as is witnessed by some recent proposals for modeling abstract concepts spatially. Here, I am going to make a further such proposal, to wit, a proposal for constructing an implicature space. I am not aware of any previous attempts to create such a space, but the idea of a conceptual space for the representation of implicatures certainly makes sense.

At least, the idea makes sense *prima facie*—there is a *concept* of conventional implicature, a *concept* of order implicature, and so on—but one must always reckon with the fact that trying to construct a conceptual space leads nowhere. To see how this may happen, it is first to be noted that conceptual spaces are typically constructed by means of a dimensionality-reduction technique, the one most commonly used being multidimensional scaling (MDS). In an MDS procedure, we construct a spatial representation of a set of items, taking as input similarity judgments, or confusion probabilities, or correlation coefficients, pertaining to those items. There is no guarantee, however, that the resulting representation will be any good. Specifically, what we aim at in an MDS procedure is a space which (i) is *low-dimensional*, ideally, with no more than three dimensions; (ii) has good fit, which in this context is expressed in terms of *stress*, where lower stress values indicate more faithful representations of the similarities/confusion probabilities/correlations related with the items we are trying to represent; and (iii) has *interpretable dimensions*, in that we can associate each dimension with some fundamental attribute the items can be said to have to some degree. An outcome of an MDS procedure may fail to satisfy some or all of these criteria.

The items we are going to use to construct an implicature space are the ones given in Table 1, and the specific input data are the correlations among the responses to those items reported in Douven and Krzyżanowska (2019) and briefly described and depicted in the previous section.

To start building our space, we must first turn those correlations into distances. There are many options for measuring such distances, but the most common ones are all instances of the so-called Minkowski metric, which is defined thus:

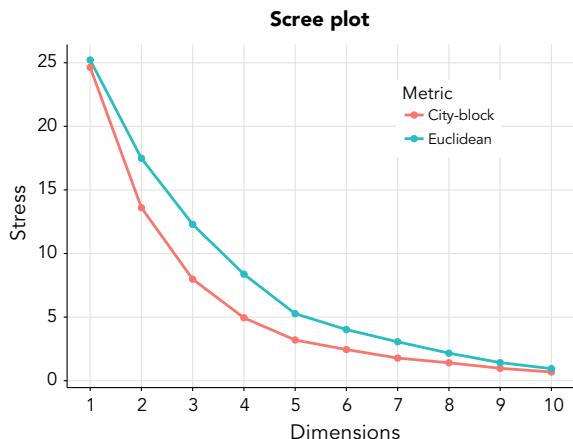
$$\delta_k(p, q) = \left( \sum_{i=1}^n |x_i - y_i|^k \right)^{1/k}$$

with  $p = \langle x_1, \dots, x_n \rangle$  and  $q = \langle y_1, \dots, y_n \rangle$ . For  $k = 1$ , this yields the so-called city-block or Manhattan metric, and for  $k = 2$ , the more familiar Euclidean metric.

It is generally held that the Euclidean metric is appropriate for measuring distances between similarity ratings (confusion probabilities, correlations) when the “dimensions” underlying those ratings are *integral* in the sense that they cannot be experienced independently of each other (for instance, one cannot separately experience the hue and the saturation of a shade). If, by contrast, the relevant dimensions are *separable* (i.e., not integral), then the city-block metric is generally considered to be the right choice (see, e.g., Torgerson 1958; Garner 1962; Shepard 1964; and Nosofsky 1986).

In the present case, it is not immediately clear which, or how many, dimensions are going to be necessary to faithfully represent our items, supposing we can obtain a faithful representation at all. Thus, in particular, it is not clear whether we should expect the dimensions to be integral or separable. For that reason, we derive distances from the correlation coefficients both via the Euclidean metric and via the city-block metric, and then carry out MDS procedures for each separately.

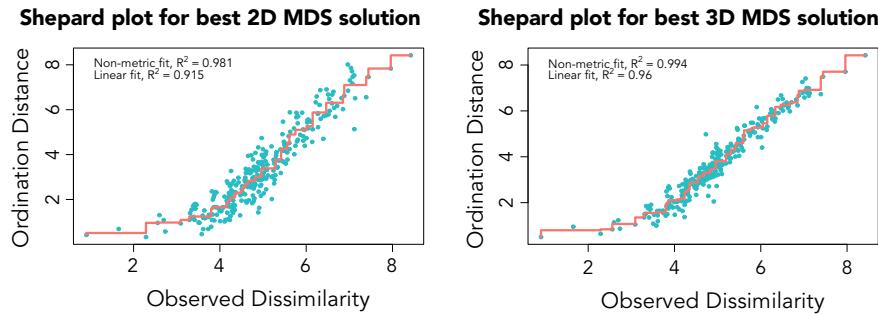
**Fig. 3** Stress values for MDS solutions with 1 to 10 dimensions, both for the city-block and for the Euclidean distances



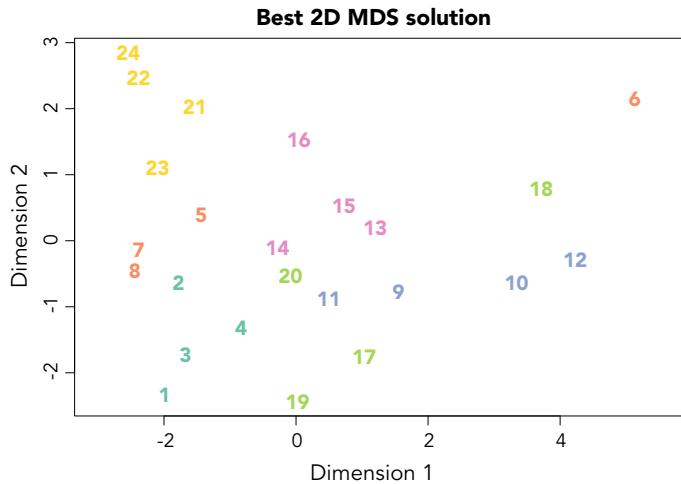
Once distances are derived—in the present case done via the `dist` function that is part of the base R language (R Core Team 2017)—one faces a further choice, to wit, whether to apply metric or nonmetric multidimensional scaling. The former tries to represent objects geometrically in a way which preserves as faithfully as possible the *distances* between those objects in the distance matrix that is given as input. By contrast, the latter tries to represent objects geometrically in a way which preserves as faithfully as possible the *ordering* of the distances between those objects according to the distance matrix; so, the smaller the distance between objects according to the matrix, the closer they are in the geometric representation, though no linear mapping of matrix distances onto distances in geometric space is aimed for. When distances derive from subjective assessments, nonmetric multidimensional scaling is generally recommended (Bartholomew et al. 2008, pp. 56–62). Given that, in our case, the distances do come from subjective assessments—people’s responses to the items in Table 1—nonmetric multidimensional scaling will be used in the following.

Specifically, we conduct the MDS procedures using the function `metaMDS` that is included in the `vegan` package for R. All configurations are centered and rotated to a principal axes orientation (see Borg and Groenen 2010, Sect. 7.10). MDS procedures are conducted for 1–10 dimensions and their stress levels are compared. The various stress values for the outcomes are shown in Fig. 3. We see immediately that we can obtain better solutions for the city-block distances than for the Euclidean distances. According to Johnson (2008, p. 205), in MDS we look for stress values less than 20. This criterion is met already by the two-dimensional solutions.

There is a second type of plot commonly used to assess the goodness-of-fit of an MDS solution, the so-called Shepard plot, in which input and output distances are plotted against each other. Figure 4 shows such plots for the best two- and three-dimensional MDS solutions, so plotting the city-block distances among the correlations (the observed dissimilarities) against the city-block distances in the solutions. We see that, in both cases, the fit is excellent, with an  $R^2$  value of .98 for the two-dimensional solution and of .99 for the three-dimensional one. Especially in the latter



**Fig. 4** Plot of distances among correlations against distances in best MDS solution

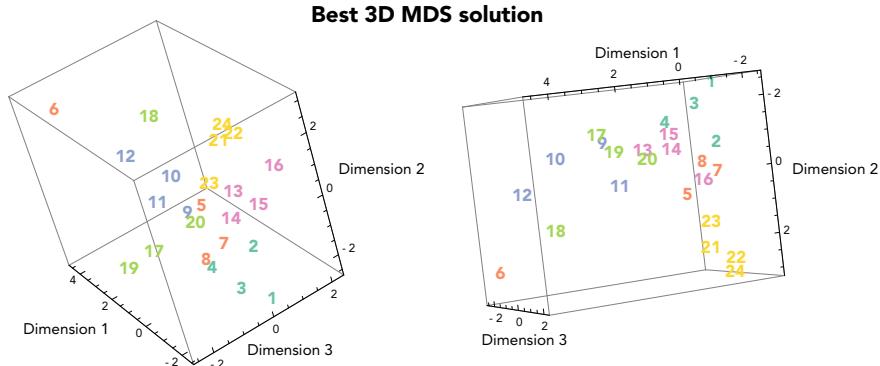


**Fig. 5** Two-dimensional MDS solution for the city-block distances; different categories of items are differently colored

case, the plotted points are grouped very tightly around the monotonically increasing line corresponding to perfect fit (for the nonmetric case). The actual solutions are displayed in Figs. 5 and 6.

So far, the best solutions satisfy two out of the three criteria (i)–(iii) mentioned above: they are low-dimensional, and they have excellent fit. How about the third criterion, that of having interpretable dimensions? While coming up with an interpretation of the dimensions of an MDS solution is often challenging (see Douven 2016a), it seems doable in the present case, at least for the first two dimensions (the only two, if we are happy to go with the two-dimensional solution).

From much of the pragmatics literature one comes away with the impression that utterances either are or are not infelicitous, depending on whether they generate a false implicature, as if that were a categorical matter. That seems as wrong, however, as the suggestion, also encountered in some of the same literature, that an utterance



**Fig. 6** Different viewpoints on the three-dimensional MDS solution for the city-block distances

either does or does not generate an implicature. The two wrong suggestions may well be related: failure to observe that utterances can be more or less felicitous may stem from a failure to observe that implicatures can be stronger or weaker.

Consider, for instance, an example from Douven (2012). In the example, a graduate student tells her supervisor,

(5) You have published some papers that I really like.

The supervisor can see two different possible explanations of why the student uttered this sentence. One is that the student wanted to convey that she read some of his papers and liked all of those; the other is that she read some or all of his papers and liked some of those she read and some not so much. The supervisor may think the first explanation tops the second and therefore infer that the student did not read all of his papers. However, the point the example is meant to illustrate is that because of the presence of an alternative explanation of why the student uttered (3), and an alternative that is close in explanation quality to the first explanation, the inference can only be guarded, so that, as a result, the implicature is only a weak one. Put differently, if it should turn out that the student read all of her supervisor's papers, an utterance of (3) would at most be minimally infelicitous.

Once this is observed, it is not too speculative to think that the first dimension represents something like degree of felicitousness (or conversely, degree of potentiality to mislead one's audience). Consider the four items most to the right in the two-dimensional space (6, 10, 12, 18), and compare them with the quantifier items (1–4) and the conventional items (21–24): All eight of the last items strike one as being much more infelicitous than the first four items. And all of the cardinal number items (13–16) do strike us as being more infelicitous than, for instance, item 6, but not quite as infelicitous as the quantifier or conventional items. More generally, that felicitousness is a matter of degree should be uncontroversial and is directly related to the claim made in Douven (2012) that implicatures can vary in strength. The latter claim was defended in terms of explanation quality—an implicature can

be part of the best explanation of why the speaker said what she said in the context in which she said it, but the extent to which the best explanation stands out as being the best can vary, and can have a significant impact on people's willingness to infer the truth of that explanation, as has recently been verified experimentally in Douven and Mirabile (2019).

Some support for this suggestion also comes from considering that item 18, which mentions Princess Diana's death first and her divorce second, carries basically *no* risk of misleading anyone about the order of the events, given that a divorce requires a person to be alive. Here, semantics (the meanings of "divorce" and "death") and world knowledge simply prohibit the implicature of the "wrong" temporal order to arise from an utterance of item 18. This is different for temporal order items 17 and 19: both suggest a temporal order of the events that is perfectly possible given the meanings of the terms involved and general knowledge about the world but that happens to be contradicted by the comic strips the sentences pertained to. Perhaps temporal order item 20 does not fit this interpretation quite as well, given that, in the context of the British royal family, it seems rather improbable, *a priori*, that the wife of a successor to the throne becomes a mother, or even becomes pregnant, while being unmarried. On the other hand, as the saying goes, the times they are a-changin'.

The strict split between conventional and conversational implicatures, mentioned in Sect. 1, may in fact be due to another false dichotomy. Against the widespread assumption that an implicature arises either due to the conventional meaning of some term or due to context plus the assumption of speaker cooperativeness, some authors have pointed out that there can be differences in the frequencies with which contexts occur that give rise to this or that implicature, and these differences may have an effect on the degree to which an implicature comes to be felt as being part of the meaning of a given expression. Hopper and Traugott (2003, Sect. 4.3) refer to this process as "semanticization," citing the following characterization of it:

[I]f some condition happens to be fulfilled frequently when a certain category is used, a stronger association may develop between the condition and the category in such a way that the condition comes to be understood as an integral part of the meaning of the category. (Dahl 1985, p. 11)

Given that the frequency with which the condition may be fulfilled in contexts in which an expression is used may vary, one would suppose that the situation that the condition is understood as part of the meaning of the expression is a limiting case, and that the strength of the association between condition and expression can vary.

To make this more concrete, compare, for instance, items 1, 8, and 22. It is difficult to imagine a context in which use of the word "therefore" does *not* suggest an inferential relationship between the clauses it connects. Helping us indicate the presence of such a relationship seems to be the *only* use we have for the word. So, it is felt as being part of the *meaning* of "therefore" that there is an inferential relationship between the connected clauses, even if for theoretical reasons it may still be better to attribute this suggestion to pragmatics—specifically, "therefore" generating a conventional implicature—than to semantics.

At the other extreme, looking at item 1, it is *very easy* to conceive of contexts in which we do not at all intend “some” to have a “not all” reading. Suppose I utter,

- (6) John is going to organize a party, and knowing him, he’s going to play loud music. Some people in the neighborhood will be annoyed.

I may utter these sentences without having any evidence, and without meaning to imply, that not all people in the neighborhood are going to be annoyed by the loud music at John’s party. What I know for sure is that *some* people are going to be annoyed, but while I am not in the stronger epistemic position to assert that *all* people are going to be annoyed, I do not wish to suggest that that is not an open possibility. And my audience, reasonably supposing that I have not surveyed all people in the neighborhood on this matter, also will not likely take me to be suggesting as much, and so will not likely be misled.

Finally, consider item 8. In virtually all contexts, we will take “somewhat cold” simply to *mean* “not extremely cold.” On the other hand, on our best current theoretical analyses of gradable adjectives (such as “cold”), these implicitly refer to standards, and such standards are known to be sensitive to contextual variation. Consider a discussion in which a group of adventurers are planning an expedition, where it is already decided that the expedition is going to be to some extremely cold place. Then the modifier “somewhat” in an utterance of item 8 might be appropriate in the context of their conversation if they had just been considering places to go where it is even colder than at the North Pole in the winter. (I am assuming, for the sake of the example, that such places exist, which I have not verified.) Even in that context, we may presume, none of the adventurers would want to deny that winter temperatures at the North Pole are extremely cold.

Perhaps similar considerations apply to the cardinal number items (13–16). Recall the context, from Sect. 1, where it would be entirely appropriate to assert that Obama has one daughter. Or consider this exchange:

Quizmaster: “Name one country that won at least four medals in the last Olympic games.”

Candidate: “France won four medals.”

Such contexts may not be very common, but they are also not extremely rare. (As for the item about Hitchcock, that may not have been well chosen, given that especially a younger generation may have little familiarity with Hitchcock or his movies.)

Based on the above considerations, and given that the conventional items are all near the bottom of the scale constituted by the second dimension, the quantifier items all at the top of that scale, and the degree modifier items as well as the cardinal number items are in between, my best guess concerning the second dimension is that it represents something like context-sensitivity or degree of semanticization.

In short, the proposed interpretations of the first two dimensions are degree of felicitousness (or degree of misleadingness) and degree of semanticization, respectively. It appears harder to come up with an interpretation of the additional dimension for the three-dimensional solution and we leave this as an open issue here. It is to be emphasized that because the MDS procedures were conducted on the basis of

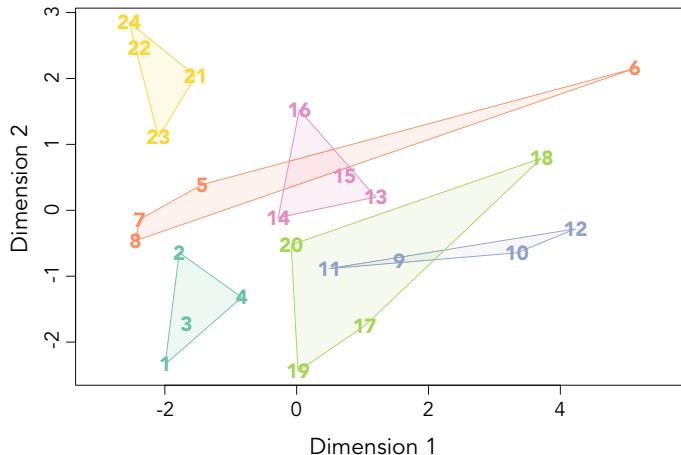
relatively sparse data, any interpretation of the dimensions is at best an exploratory hypothesis, to be confirmed in follow-up research, ideally involving a richer set of materials.

## 4 Naturalness

We finally come to the question concerning naturalness: Are the concepts associated with the various types of implicatures *natural* ones? We did much of the necessary stage-setting in the previous section, due to which we now have available an implicature space (or two, if we like), which will make answering the aforementioned question much easier. After all, as was remarked in Sect. 1, in the conceptual spaces framework the notion of naturalness has a precise meaning, or at least the framework provides a precise criterion for naturalness, viz., convexity. (It will be recalled that a region is convex if and only if, for any pair of points lying in the region, the line segment connecting them lies in its entirety in the region as well.) As mentioned, there is a wealth of evidence supporting this criterion; for instance, in color space, we find only shades of red between any pair of shades of red, and not also (say) shades of blue or green or orange. Does a similar conclusion hold for the various types of implicatures as represented in our implicature space(s)?

We start by considering again the two-dimensional solution shown in Fig. 5. We observe that, in this solution, the quantifier items (1–4) are tightly grouped together, as are the cardinal number items (13–16) and the conventional items (21–24). The same is true for three of the four gradable adjective items (5, 7, 8), the outlier being 6. One reason why this may not be very surprising is that the first three items all concern so-called degree modifier phrases (“*X* is relatively/moderately/somewhat *Y*”), whereas the outlier involves a comparison class phrase (“*X* is *Y* for a *Z*”). In the pragmatics literature, these are commonly distinguished, and so it might have been better if Douven and Krzyżanowska had kept them separate in their work; they might for instance have included four items of each subtype among their materials. In any case, the types seem to trigger somewhat different pragmatic inferential mechanisms: degree modifier phrases implicate that the utterance would be false, or at least further from the truth, were the modifier omitted, while comparison class phrases implicate that the utterance would be false, or at least further from the truth, if the comparison class were not mentioned or were replaced by the normally implicit default comparison class (“Trump is rich for an American president” implicates that he is not rich *tout court*, or not rich for an American, generally speaking).

There may be an even simpler explanation for the outlier. The assertion that Margo Dydek was tall for a woman will normally generate the implicature that she is *not* tall for a person (when the men are included in the comparison class), which is false in the present case. But while thereby an assertion of item 6 would *normally* generate a false implicature, and so would *normally* be misleading, Douven and Krzyżanowska could only assume their participants to see the falsity of the implicature by adding, in parentheses, the height of Margo Dydek (everybody knows Bill Gates, and knows

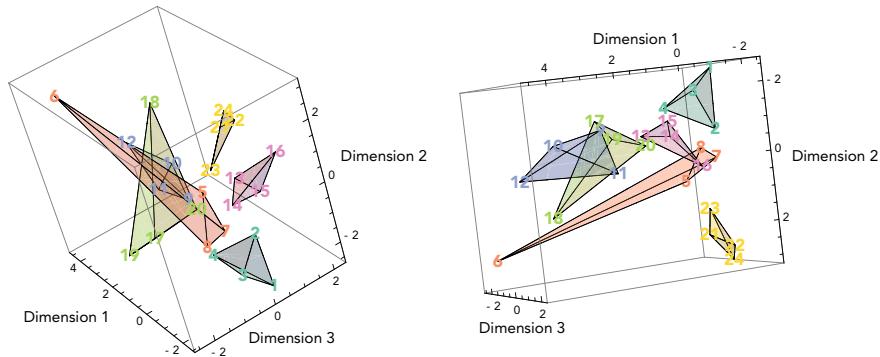


**Fig. 7** Two-dimensional MDS solution with convex hulls added

that he is rich, but not so many will have heard of Margo Dydek). However, with the basketball player's height being explicitly mentioned in the sentence, even if only parenthetically, the risk of generating a false implicature is automatically reduced to zero: the sentence, while somewhat awkwardly formulated perhaps, will have no tendency to mislead anyone into thinking that Margo Dydek was not tall for a person (being over 2 m, as the sentence asserts her height is, counts as tall by any reasonable standard). In retrospect, then, this was probably a poorly chosen item in Douven and Krzyżanowska's materials.

At first blush, the picture appears to be more troubling for the ranked ordering items (9–12) and the temporal order items (17–20). In neither group do the items seem to hang together very tightly. More importantly still, they do not appear to form convex regions in the space. Whereas, as just mentioned, we do not find shades of blue or green among the shades of red in color space, from Fig. 5 it looks as though the ranked ordering items and the temporal order items *are* interspersed. (The fact that both types refer to some kind of *ordering* could lead one to believe that maybe these items form actually only one type of implicature, which might then be represented by a convex region. But that would be a mistake: the orderings have nothing essentially in common, ranked ordering implicatures implicitly referring to some scale, and temporal ordering implicatures explicitly referring to different points in time, even if the points in time can remain unspecified.) This becomes easier to see still when we add, as is done in Fig. 7, the convex hulls for the different types of implicatures to the MDS solution. (The convex hull of a set of points is the smallest convex set encompassing all points in the set.)

The three-dimensional MDS solution scored better on stress than the two-dimensional one, and it might be that all types of implicatures do form convex regions in three-dimensional space. This is *almost* the case, but here, too, the ranked ordering items and the temporal order items have partly overlapping convex hulls (all



**Fig. 8** Three-dimensional MDS solution with convex hulls added

other convex hulls are cleanly separated from each other). This can be seen *somewhat* from Fig. 8, although it is only really clear if one rotates the figure in *Mathematica*, the software that was used to produce the plots.

So, we might be inclined to conclude that either ranked ordering implicatures or temporal order implicatures (or both) fail to constitute a natural concept, or at any rate not one as natural as the other types of implicatures. I doubt, however, whether that conclusion would be warranted. Specifically, I doubt whether we should assume that all alleged ranked ordering items and all temporal order items in Douven and Krzyżanowska's materials generate the implicatures they were supposed to generate.

When considering an interpretation of the first dimension of the implicature space (or spaces), we already noted that some conjunctions that relate events in the wrong temporal order will nonetheless not lead hearers to make any false inferences about that order. That is simply because some events can only occur in a given order, for logical reasons, or probably more often for reasons of how the world is organized, whether physically, biologically, legally, socially, or in some other respect. In particular, item 18, about Princess Diana, will not have led anyone to believe, even if only for a moment, that she first died in a car accident and then had a divorce. And rerunning the whole MDS procedures described in the previous section but now leaving item 18 out does produce a space in which all types of implicatures form convex concepts.

This is not necessarily to say we should put all the blame on item 18. Some of the ranked ordering items may not have been as happily chosen either. For instance, it is conceivable that item 12, about Americans earning over \$200,000 a year having to pay taxes, may for some of Douven and Krzyżanowska's participants not even have generated a weak implicature to the effect that Americans earning less are exempt from paying taxes. That is because the item is easily interpretable as making an assertion about a specific income group with no intention to suggest anything about any other income group, or so it seems.

More generally, at this point it is probably best not to make too much of the apparent clash of the temporal order and ranked ordering implicatures in the two-

and three-dimensional spaces, and rather to take the finding as motivating further research, with a richer set of materials, which is at the same time better geared to the specific purpose of constructing a conceptual space.

## 5 Concluding Remarks

The main question addressed in this paper was whether the various types of implicatures postulated by modern-day pragmatics constitute natural concepts. The question is an important one insofar as serious scientific theories are supposed to feature precisely such concepts. To answer this question, some preparatory work had to be done, mainly in the form of constructing an implicature space. We followed a common procedure for constructing such spaces, noting however that there was no guarantee that the procedure would work. But we were lucky and ended up with a two-dimensional implicature space that met all criteria by which conceptual spaces are commonly judged. We also obtained a three-dimensional space that appeared to fit the input data even better, although here we had some difficulty interpreting all three dimensions (a problem that might be overcome by gathering further data and rerunning the analysis).

Examination of where in our space (or spaces) the items that had served as input were located showed a tight within-type clustering of most of those items. More importantly still, items belonging to the same type tended to span convex regions in that items belonging to one type lay mostly not between items belonging to some other type. While this is not *proof* that the various types of implicatures correspond to natural concepts—given that convexity is only a necessary criterion—it is at least some first evidence that they do correspond to such concepts indeed.

Admittedly, there were some violations of the convexity criterion. The results might in fact lead one to speculate that temporal order implicatures do not constitute a natural class, or not a highly natural one (if naturalness comes in degrees). One might even be able to back this speculation up theoretically, by pointing out that there may not be a one-to-one relation between respecting temporal order in a sentence and risk of misleading one's audience by uttering that sentence, given that the latter may be prevented by world knowledge even if the sentence relates events in the wrong order. But, as noted in the previous section, this speculation is probably best not taken too seriously at the moment, given that our results were based on relatively sparse materials, which on top of that were not chosen with an MDS-kind of analysis in mind.

What we have, then, is a proof of principle that implicatures can be represented in a conceptual space, and that this can help answering an important theoretical question about them. That is good news for researchers interested in experimental pragmatics, as conceptual spaces make it easy to generate empirical predictions about which factors will determine the classification of whichever items are representable in them. And it is equally good news for advocates of the conceptual spaces framework, who are constantly looking for ways to generalize their framework to domains beyond

those of perceptual concepts. But to see exactly how much research on implicatures can benefit from the current approach, more empirical work is called for, along the lines hinted at at various junctures in this paper.<sup>2</sup>

## References

- Bach, K. (1999). The myth of conventional implicature. *Linguistics and Philosophy*, 22, 327–366.
- Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. I. (2008). *Analysis of multivariate social science data* (2nd ed.). Boca Raton FL: CRC Press.
- Borg, I. & Groenen, P. (2010). *Modern multidimensional scaling* (2nd ed.). New York: Springer.
- Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies*, 61, 127–148.
- Breheny, R. (2008). A new look at the semantics and pragmatics of numerical quantified noun phrases. *Journal of Semantics*, 25, 93–140.
- Dahl, Ö. (1985). *Tense and aspect systems*. Oxford: Blackwell.
- Decock, L., & Douven, I. (2014). What is graded membership? *Noûs*, 48, 653–682.
- Douven, I. (2010). The pragmatics of belief. *Journal of Pragmatics*, 42, 35–47.
- Douven, I. (2012). The lottery paradox and the pragmatics of belief. *Dialectica*, 66, 351–373.
- Douven, I. (2016a). Vagueness, graded membership, and conceptual spaces. *Cognition*, 151, 80–95.
- Douven, I. (2016b). *The epistemology of indicative conditionals*. Cambridge: Cambridge University Press.
- Douven, I. (2018). The rationality of vagueness. In R. Dietz (Ed.), *Vagueness and rationality*. New York: Springer, in press.
- Douven, I., & Decock, L. (2017). What verities may be. *Mind*, 126, 386–428.
- Douven, I., Decock, L., Dietz, R., & Égré, P. (2013). Vagueness: A conceptual spaces approach. *Journal of Philosophical Logic*, 42, 137–160.
- Douven, I., & Gärdenfors, P. (2018). What are natural concepts? A design perspective, manuscript.
- Douven, I., & Krzyżanowska, K. H. (2019). The semantics-pragmatics interface: An empirical investigation. In A. Capone, M. Carapezza, and F. Lo Piparo (Eds.), *Further advances in pragmatics and philosophy: Theories and Applications* (pp. 81–100) (Vol. II). New York: Springer, in press.
- Douven, I., & Mirabile, P. (2018). Best, second-best, and good-enough explanations: How they matter to reasoning. *Journal of Experimental Psychology: Memory, Learning, and Cognition*, in press.
- Douven, I., & van Brakel, J. (1998). Can the world help us in fixing the reference of natural kind terms? *Journal for General Philosophy of Science*, 29, 59–70.
- Douven, I., Wenmackers, S., Jraissati, Y., & Decock, L. (2017). Measuring graded membership: The case of color. *Cognitive Science*, 41, 686–722.
- Fairchild, M. D. (2013). *Color appearance models*. Hoboken NJ: Wiley.
- Gärdenfors, P. (2000). *Conceptual spaces*. Cambridge MA: MIT Press.
- Gärdenfors, P. (2018). Convexity is a testable criterion in the theory of conceptual spaces. In M. Kaipainen, F. Zenker, A. Hautamäki, & P. Gärdenfors (Eds.), *Conceptual spaces: Elaborations and applications*. New York: Springer, in press.
- Garner, W. R. (1962). *Uncertainty and structure as psychological concepts*. New York: Wiley.
- Grice, H. P. (1989). Logic and conversation. In M. A. Cambridge (Ed.), *His studies in the ways of words* (pp. 22–40). Harvard University Press.
- Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization* (2nd ed.). Cambridge: Cambridge University Press.

---

<sup>2</sup>I am greatly indebted to two anonymous referees for helpful comments on a previous version.

- Jraissati, Y., & Douven, I. (2018). Delving deeper into color space, manuscript.
- Johnson, K. (2008). *Quantitative methods in linguistics*. Oxford: Blackwell.
- Lewis, D. K. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, 61, 343–377.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, 115, 39–57.
- Potts, C. (2005). *The logic of conventional implicatures*. Oxford: Oxford University Press.
- Psillos, S. (1999). *Scientific realism*. London: Routledge.
- Putnam, H. (1983). *Realism and reason*. Cambridge: Cambridge University Press.
- R Core Team (2017) R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, <http://www.R-project.org/>.
- Scharten, R. (1997). *Exhaustive interpretation: A discourse-semantic account*, Ph.D. dissertation, University of Nijmegen.
- Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, 1, 54–87.
- Spychalska, M., Kontinen, J., & Werning, M. (2016). Investigating scalar implicatures in a truth-value judgement task: Evidence from event-related brain potentials. *Language, Cognition and Neuroscience*, 31, 817–840.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Perception, Types and Frames



Robin Cooper

**Abstract** We present a view of perception as the classification of objects and events in terms of types in the sense of TTR, a Type Theory with Records. We argue that such types can be used to give a formal model of concepts and cognitive processing involving concepts. This yields a view that natural language semantics is based on our cognitive perceptual ability. The paper provides an overview of some key ideas in TTR including the important notion of record type. We suggest that record types can be used to model frames in a way that relates to the Düsseldorf notion of frame as well as those of Fillmore and Barsalou.

**Keywords** Frames · Record types · Partee puzzle · Coercion

## 1 Introduction

We will present a simple-minded view of perception as the classification of objects and events in terms of types viewed as cognitive resources. The theory of types that we are using is TTR, a Type Theory with Records, which borrows a great deal from work in logic and computer science in a tradition initiated by Per Martin-Löf. It provides a rich type theory, that is, it includes types not just for basic ontological categories such as entities and functions, but also types of objects such as *Tree* and *Boy* and types of events (or situations) such as *Hugging-of-a-dog-by-a-boy*. Types may be complex objects constructed from other types in a type theoretic universe. We will argue that such types can be used to give a formal model of concepts and cognitive processing involving concepts. In particular, we will suggest that natural language semantics is at bottom based on our cognitive ability to perceive objects and situations in terms of types. To this we have added the ability to reason in terms of

---

R. Cooper (✉)  
University of Gothenburg, Gothenburg, Sweden  
e-mail: [cooper@ling.gu.se](mailto:cooper@ling.gu.se)

the types themselves. Thus, for example, we can consider types of situations without actually perceiving a situation of the type and we can even consider types of situations which are impossible.

Among the complex types introduced in TTR are record types which are used to model types of situations and also propositions. An utterance of the sentence *A boy hugged a dog* is true if there is a situation of the type *Hugging-of-a-dog-by-a-boy* and false if there is nothing of that type. (This follows the dictum known as “Propositions as Types” which Martin-Löf took over from intuitionistic logic.) Both the intuition behind record types and their structure in the formal theory suggest that they can be used to model frames, both as conceived of by Fillmore and as introduced by Barsalou. We will develop this correspondence and suggest that this provides one way of integrating frames into compositional semantics. In exploring this we will find relations with work on frames conducted by several researchers in the Düsseldorf group working on frames.

## 2 Types and Cognition

Here we will give a brief overview of certain key ideas in TTR. For more detailed discussion of TTR in general see Cooper (2012, prep), Cooper and Ginzburg (2015). TTR is a *rich* type theory: in contrast to the simple type theory used in formal semantics as developed by Montague (1974), it contains a much richer collection of types. Whereas Montague has types for what we might call basic ontological categories such as entities and truth values, TTR includes types of objects like *Tree* and of events such as *boy-hugs-dog*. We will see later that such types may have a complex internal structure. For discussion of the difference between simple and rich type theories including a historical perspective see Chatzikyriakidis and Cooper (2018). TTR is inspired by work in the tradition of Martin-Löf type theory (Martin-Löf 1984; Nordström et al. 1990). While it has borrowed many tools and insights from this it does not follow all of the basic tenets of Martin-Löf type theory such as a proof-theoretic constructive approach derived from intuitionism. For discussion of some of the differences and motivations see Cooper (2017a).

A central notion in Martin-Löf type theory is *judgement*, a judgement that an object (or event),  $a$ , is of type,  $T$ . This is represented in symbols in (1).

$$(1) \quad a : T$$

We say that  $a$  is a *witness* for  $T$ . In work using TTR we put a cognitive spin on this notion. Suppose an agent,  $A$ , perceives a tree,  $t$ . (Here we are thinking of  $t$  as an object in the world, construed naively, that is the physical object with a trunk, branches and leaves.) We say that perception involves classifying an object as being of some particular type, that is making a judgement. Thus perceiving  $t$  as a tree,  $A$  makes the judgement that  $t$  is of type *Tree*. In symbols we can write this as (2).

(2)  $t :_A Tree$

For discussion of this notation and the theory of type acts that we associate it with see Cooper (2014). We can think of the type *Tree* as what Gibson (1979) would call an *invariant*: whatever it is that trees share in common that enable us to classify them as trees. Following Gibson’s terminology we can say that  $A$  is *attuned* to this type or  $A$  has this type as a *resource*. The idea that attunement is an important notion for semantics goes back to work on situation semantics (Barwise and Perry 1983), which is another important source of inspiration for our work on TTR.

Different agents have different type resources available. For example, a bee landing on the tree perceived by  $A$  probably does not have the same type *Tree* as the human  $A$  does. Different species have different perceptual apparatus and cognitive abilities. Even within a species the resources we have available might vary depending on our experience. For example, most people have a greater variety of subtypes for *Tree* than I do corresponding to different kinds of trees. The idea of linking types to perception is developed further by Larsson (2013) and is related to the theories which ground cognition in perception, for example, Barsalou (1999). For an agent to be able to make classifications corresponding to types there must be patterns of neural activation corresponding to types which we could think of as *mental representations* of types. For some suggestions concerning how such neural representations might be see Cooper (2017b, 2019).

TTR provides not only types of objects but also types of situations, following a suggestion by Ranta (1994). Suppose that the boy, Sam, hugs his dog, Fido. The type of situation in which Sam hugs Fido is represented as in (3).

(3)  $\text{hug}(\text{sam}, \text{fido})$

We are used to this notation as a logical formula which denotes a truth value. In TTR, however, we use the notation to represent a type of situation. Nevertheless we can recover the notion of truth by using the “propositions as types” dictum (see Chatzikyriakidis and Cooper 2018 for discussion and references). A type (thought of as a proposition) is true just in case it has a witness, that is, there is something of the type. The type (3) is a complex type which is constructed from the predicate ‘hug’ and two individuals (‘sam’ and ‘fido’) as arguments.

Suppose, however, that we want a more general type of situation, one where any boy hugs any dog, that is, the type *Hugging-of-a-dog-by-a-boy* which we mentioned in Sect. 1. In TTR we use *record types* for this. Consider the record type in (4).

(4)

$$\left[ \begin{array}{l} x : Ind \\ c_{\text{boy}} : \text{boy}(x) \\ y : Ind \\ c_{\text{dog}} : \text{dog}(y) \\ e : \text{hug}(x,y) \end{array} \right]$$

This is a graphical notation for a set of *fields*, which in turn are ordered pairs containing a *label* and a type. The type *Ind* is the type of individuals, about which we say more below. A type like ‘boy( $x$ )’ is a *dependent type*—exactly which type it

is depends on the individual you choose in the ‘x’-field. A witness for this record type is also a set of fields, though in this case the fields consist of a label followed by an object. A record is a witness for the record type if it contains fields with the same labels as the type (and possibly more fields with other labels) and the objects in these fields are witnesses for the corresponding types in the record type. So, for example, a record of the form (5a) would be a witness for (4) provided that it meets the conditions in (5b).

- (5) a.
- $$\left[ \begin{array}{l} x = \text{sam} \\ c_{\text{boy}} = s_1 \\ y = \text{fido} \\ c_{\text{dog}} = s_2 \\ e = s_3 \\ \dots \end{array} \right]$$
- b.  $\text{sam} : Ind$   
 $s_1 : \text{boy}(\text{sam})$   
 $\text{fido} : Ind$   
 $s_2 : \text{dog}(\text{fido})$   
 $s_3 : \text{hug}(\text{sam}, \text{fido})$

We can think of records as modelling complex situations in which each field introduces either an object or a situation. Thus we can think of (4) as being the type of situations in which a boy hugs a dog.

What does it mean for an agent to perceive some situation,  $s$ , as being of type (4)? If situations are to be construed as being part of the world (as in Barwise and Perry 1983) then we might be misled by thinking of a situation as being of the type (4). After all (4) is a record type and a record, as we have seen, is a pairing of labels with objects like Sam and situations in which, for example, Sam is a boy or, if you like, *proof objects*, such as a part of the world which shows that Sam is a boy. (The term *proof object* was introduced by Martin-Löf and shows an important bridge between a proof theoretic and a model theoretic approach to logic.) While it seems reasonable (though not entirely uncontroversial) to say that objects like Sam and situations in which he is a boy are parts of the world, the world does not come conveniently labelled as would be suggested by a record. We do not wish to claim that the world consists of records as characterized in TTR. The notation (6a) in TTR is a convenient graphic display of a set of ordered pairs (the graph of a function) whose first members are labels and whose second members are objects, as in (6b).

- (6) a.
- $$\left[ \begin{array}{l} x = \text{sam} \\ c_{\text{boy}} = s_1 \\ y = \text{fido} \\ c_{\text{dog}} = s_2 \\ e = s_3 \end{array} \right]$$
- b.  $\{\langle x, \text{sam} \rangle, \langle c_{\text{boy}}, s_1 \rangle, \langle y, \text{fido} \rangle, \langle c_{\text{dog}}, s_2 \rangle, \langle e, s_3 \rangle\}$

Another intuitive way to think about this is as a labelling of the set<sup>1</sup> (7a) which could be graphically represented as (7b).

- (7) a. {sam,  $s_1$ , fido,  $s_2$ ,  $s_3$ }

b.	x	$c_{\text{boy}}$	y	$c_{\text{dog}}$	e	

{sam,  $s_1$ , fido,  $s_2$ ,  $s_3$ }

Intuitively, the elements in the set in (7b) are part of the world whereas the labels are pointers or handles introduced by cognitive processing of the world. Depending on your metaphysical view, you can consider the set in (7a), as opposed to the elements of the set either as something existing in the world or a cognitive construct which assembles those elements into a collection. On our view, records, at any rate, represent cognitive objects since they introduce labelling and perception of a situation as one in which a boy hugs a dog involves breaking down the situation into components corresponding to the boy, the dog, the “boyness” of the boy, the “dogginess” of the dog and the “hugging” event involving the boy and the dog.

It might be that we could regard this as perception of a collection of tropes according to one or more of the varieties of tropes that have been proposed (Maurin 2016).<sup>2</sup> A witness for a type like ‘boy(sam)’ is normally glossed in TTR as a situation which shows (or proves) that ‘sam’ is a boy. Such a situation is a particular (an “object” in TTR terms) as required for a trope though it is perhaps not clear that it is abstract in the right sense for a trope. It appears, at any rate, that it would not be the kind of trope discussed by Moltmann (2013). For one thing, Moltmann does not consider tropes as corresponding to common nouns in natural language. For another, there seems to be a kind of uniqueness of tropes instantiated by particular objects as in *the red of the box* whereas on our view given a box  $b$ , there could be many witnesses for the type ‘red( $b$ )’, that is, situations which are proofs for the redness of the box. Furthermore the red of the box would be shared with another box which has exactly the same shade of red. There is no requirement that a situation which shows that one box is red also shows another box to be red, although there can be such situations. However, a situation which shows two boxes to be red would not require that the two boxes have an identical shade of red. This would, in Moltmann’s terms at least, indicate that the situation is not a trope. Nevertheless, there is something trope-like about the situations which witness these types in that they are particulars which instantiate a specific quality obtained by applying a single predicate to appropriate arguments.

Record types give us a notion of *subtyping*. We can obtain a subtype of a record type by adding additional fields to it. Any record of the type with additional fields will also be of the type with fewer fields because a witness for a record type may contain additional fields with labels not occurring in any field in the record type. Thus the intuitive fact that any situation in which a boy hugs a dog is a situation in which there is a boy is modelled by the subtype relation expressed in (8).

---

<sup>1</sup>In general, records correspond to multisets since objects may occur more than once in a record.

<sup>2</sup>I am grateful to one of the anonymous referees for raising this possibility.

$$(8) \quad \left[ \begin{array}{ll} x & : Ind \\ c_{\text{boy}} & : \text{boy}(x) \\ y & : Ind \\ c_{\text{dog}} & : \text{dog}(y) \\ e & : \text{hug}(x,y) \end{array} \right] \sqsubseteq \left[ \begin{array}{ll} x & : Ind \\ c_{\text{boy}} & : \text{boy}(x) \end{array} \right]$$

We have talked as if there are situations like a boy hugging a dog on the one hand and objects like trees on the other, but actually the dividing line between them is not so obvious. For example, you could think of *Tree* as being shorthand for a record type like (9).

$$(9) \quad \left[ \begin{array}{ll} x & : Ind \\ y & : \text{set}(Ind) \\ c_{\text{leaves}} & : \text{leaves}(y,x) \\ z & : \text{set}(Ind) \\ c_{\text{branches}} & : \text{branches}(z,x) \\ w & : Ind \\ c_{\text{trunk}} & : \text{trunk}(w,x) \end{array} \right]$$

(Here ‘ $\text{set}(Ind)$ ’ represents the type of sets of individuals.) This represents the intuition that trees have leaves, branches and a trunk. You can either think of this as an individual or as a situation in which various things hold. Using the type *Ind* for “individual” as we standardly do in TTR, following the lead of traditional model theoretic semantics (*cf.* Montague’s type *e*), hides a great deal of complexity which needs attention if we are to take a cognitive approach to perception and semantics. Perhaps the least you can say is that each agent may have their own view of what counts as a witness for *Ind* corresponding to a scheme of individuation (discussed in connection with semantics by, for example, Barwise 1989). For important work addressing some of the many difficulties involving individuation see Sutton and Filip (2017).

In this section we have talked about types from a cognitive perspective and in fact we can think of types as models of cognitive notions like concept, memory and belief. If we think of a concept as a type we can say that the concept is instantiated just in case there is a witness of the type. If we think of a memory as a type we can say that the memory is correct just in case there is or was a witness for the type. If we think of a belief as a type we can say that the belief is true just in case there is a witness for the type. This, coupled with the ideas of how types could be represented on a network of neurons presented by Cooper (2017b, 2019), gives us an admittedly very preliminary and “armchairish” theory of how concepts, memories and beliefs could be represented in the brain. It is my hope that this might in the future lead to a substantial connection between formal work on language and empirically based neuroscience. It is in this context that I would like to view the discussion of frames in the next section.

### 3 Record Types and Frames

TTR has been used to model frames by Cooper (2010, 2016). This work took the frame semantics suggested by Fillmore (1982, 1985) leading to the kind of frames used in FrameNet (<https://framenet.icsi.berkeley.edu>) as its starting point. However, the use of frames to analyze the Partee temperature puzzle is strikingly similar to that proposed by Löbner (2014, 2015) who based his work on Barsalou's (1992) more cognitively based notion of frame.

Partee's temperature puzzle involves explaining why the inference in (10) is not valid, as it would be if the interpretation of *is 90* is "is identical with 90".

- (10) The temperature is 90  
The temperature is rising  
 90 is rising

In order to address this puzzle Cooper (2016) uses the record type (11) corresponding to a stripped down version of the FrameNet frame `Ambient_temperature`.

$$(11) \quad \left[ \begin{array}{l} x : \text{Real} \\ \text{loc} : \text{Loc} \\ e : \text{temp(loc, x)} \end{array} \right]$$

We call (11) *AmbTempFrame*. Any record belonging to this type will contain a pair of a real number (in the 'x'-field) and a location (in the 'loc'-field) such that the real number is the temperature at the location. In the terminology adopted in Cooper (2016) we refer to the record type *AmbTempFrame* as a frame type and we refer to records that are witnesses for it as frames. As records are used to model situations (including both states and events) frames correspond to situations and frame-types correspond to situation types. The basic idea in Cooper (2010, 2016) is that a temperature rise is a string of two frames,  $s_1 s_2$ , such that  $s_1, s_2 : \text{AmbTempFrame}$  and  $s_1.\text{loc} = s_2.\text{loc}$  and  $s_1.x < s_2.x$ . This is a very simple theory of temperature rises. One might, for example, object to holding the location constant in view of sentences like (12).

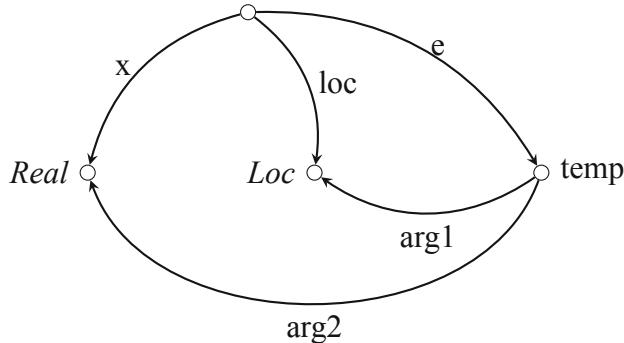
- (12) The temperature rises as you go south

Cooper (2016) suggests, however, that all locations are relative, even those we consider to be fixed locations on the Earth when we consider them from an astronomical perspective, so we could think of the location in (12) as being the relative location "around you". One might object also to having a string of just two frames corresponding intuitively to two temperature readings over time. The idea of strings is adapted from Fernando's (2004, 2006, 2008, 2009, 2011, 2015) work on a string theory of events, where a finite string can be regarded as a finite number of observations of a continuous world. The question arises whether the temperature should be rising between the two frames or whether it would still count as a rise even if the temperature was lower at some point between the two frames. The fact that examples like (13) can be true despite temperature dips during the night suggests that we can allow for temperature falls during a rise.

- (13) The temperature rose during the week

*AmbTempFrame* can be related to a directed graph similar to those discussed by Kallmeyer and Osswald (2013), Kallmeyer et al. (2017) in connection with frames. We let the labels in the record type be labels on the edges and the types be labels on the nodes. In the case of types constructed with a predicate we use the predicate to label a node with edges labelled ‘argn’ corresponding to the arguments of the predicate. Thus the type *Ambient\_temperature* in (11) could correspond to the directed graph in (14).

- (14)



This would indicate that ambient temperature has three attributes: a real number (here labelled as the attribute ‘x’), a location and a constraint (here labelled as the attribute ‘e’) that the real number is the temperature at the location.

Both the record type (11) and the directed graph (14) could be coded in terms of hybrid logic in the manner suggested in Kallmeyer et al. (2017) as in (15).

- (15)  $\langle x \rangle(l_1 \wedge \text{Real}) \wedge \langle \text{loc} \rangle(l_2 \wedge \text{Loc}) \wedge \langle e \rangle(\text{temp} \wedge \langle \text{arg1} \rangle l_2 \wedge \langle \text{arg2} \rangle l_1)$

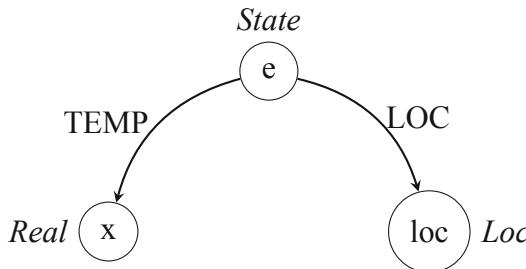
One of the anonymous referees offers a different way of relating TTR frames and Düsseldorf frames (DF). This involves thinking of the attributes in DF as functions from entities to entities in TTR. What appears below is my own adaptation of the referee’s suggestion and the referee (anonymous, though he or she is) should not be held responsible for it. The suggestion involves first recasting the TTR frame type suggested in (11) in a neo-Davidsonian version, something that I think can be a good idea in many respects although it has not been explored to any extent within TTR. My suggestion for a neo-Davidsonian type for ambient temperature is given in (16).

$$(16) \quad \left[ \begin{array}{l} e : \text{State} \\ x : \text{Real} \\ \text{loc} : \text{Loc} \\ c_1 : \text{LOC}(e, \text{loc}) \\ c_2 : \text{TEMP}(e, x) \end{array} \right]$$

The referee’s idea is that then the labels in fields with basic types stand in for values in DF and the predicates in the types labelled ‘c<sub>1</sub>’ and ‘c<sub>2</sub>’ correspond to attributes

which label edges in the DF graph. Thus we would obtain (17), again a modification of the reviewer's original.

(17)



This certainly gives us a more intuitive looking Düsseldorf frame. Also the representation of this in hybrid logic, given in (18), corresponds more closely to the use of hybrid logic by Kallmeyer et al. (2017).

(18)  $e \wedge \text{State} \wedge \langle \text{TEMP} \rangle (x \wedge \text{Real}) \wedge \langle \text{LOC} \rangle (\text{loc} \wedge \text{Loc})$

A possible disadvantage with this, though, is that the relationship between the record type and the directed graph is less direct than in the first suggestion that we presented.

This discussion raises the interesting question of whether a general relationship could be shown between TTR and hybrid logic and more specifically between frames modelled in terms of records and record types and frames as modelled by Kallmeyer and Osswald (2013), Kallmeyer et al. (2017). Then it is interesting to consider whether the particular linguistic analyses offered in the two approaches to frames can be intuitively represented in both TTR and DF.

For example, it is not obvious to me that the following analysis could be easily reconstructed in DF, although I would be happy to be convinced otherwise. The basic idea in Cooper (2010, 2016), although the analyses in the two papers differ in details, is that *temperature* and *rise* correspond to predicates not of numbers but of frames of the type *AmbTempFrame* and for this reason the offending inference in the Partee puzzle does not go through. This leads us to distinguish between nouns and verbs which correspond to properties of individuals on the one hand and properties of frames on the other. The way that this distinction is made in Cooper (2016) is represented in (19) where *dog* and *run* correspond to individual level properties and *temperature* and *rise* frame level properties (modelled as properties of records).

- (19) a.  $\text{dog} — \lambda r: [\text{x:Ind}] . [e : \text{dog}(r.x)]$   
 b.  $\text{temperature} — \lambda r: [\text{x:Rec}] . [e : \text{temperature}(r.x)]$   
 c.  $\text{run} — \lambda r: [\text{x:Ind}] . [e : \text{run}(r.x)]$   
 d.  $\text{rise} — \lambda r: [\text{x:Rec}] . [e : \text{rise}(r.x)]$

However, things are not quite so straightforward. Consider the putative inference in (20) which apparently is an instance of the Partee puzzle involving individual level properties.

- (20) The dog is nine  
The dog is getting older  
 Nine is getting older

The conclusion drawn by Cooper (2016) is that expressions corresponding to individual level properties can have a coerced interpretation where they correspond to frame level properties.<sup>3</sup> Thus in addition to (19a) we can obtain a coerced interpretation of *dog* as in (21).

- (21)  $\lambda r:[x:Rec]. [e : \text{dog\_frame}(r.x)]$

A record is a dog frame just in case it is of the type (22a). For example, it may be of the type (22b), a subtype of (22a).

- (22) a.  $\begin{bmatrix} x:Ind \\ e:\text{dog}(x) \end{bmatrix}$
- b.  $\begin{bmatrix} x:Ind \\ e:\text{dog}(x) \\ \text{age:Real} \\ c_{\text{age}}:\text{age\_of}(x,\text{age}) \end{bmatrix}$

This allows for frames of types other than (22b) to count as dog frames. The only requirement on a dog frame is that it contain an individual which is a dog. What other information we put into the frame may vary with whatever we are interested in when creating the frame. For many objects age is a relevant issue and we can imagine that among our resources is the type (23a) (which requires an individual with some age) and that this type can be merged with a minimal frame type like (22a) as indicated in (23b).

- (23) a.  $\begin{bmatrix} x:Ind \\ \text{age:Real} \\ c_{\text{age}}:\text{age\_of}(x,\text{age}) \end{bmatrix}$
- b.  $\begin{bmatrix} x:Ind \\ e:\text{dog}(x) \end{bmatrix} \wedge \begin{bmatrix} x:Ind \\ \text{age:Real} \\ c_{\text{age}}:\text{age\_of}(x,\text{age}) \end{bmatrix} = \begin{bmatrix} x:Ind \\ e:\text{dog}(x) \\ \text{age:Real} \\ c_{\text{age}}:\text{age\_of}(x,\text{age}) \end{bmatrix}$

(For the notion of merge in TTR (represented by ‘ $\wedge$ ’ see discussion in Cooper 2019; Cooper and Ginzburg 2015.) Thus (23a) could be thought of as a resource which could be used in a general coercion procedure for taking individual level properties to frame level properties involving a frame type including age information.

This, perhaps, points to a rather different notion of frame than we have in either Fillmore’s or Barsalou’s work where we get the impression that frames might be a fixed non-dynamic part of our cognitive furniture. This appears to be the case

---

<sup>3</sup>The terminology “individual/frame level” here is meant to suggest a parallel with the well-known distinction which is drawn between individual, stage and kind level predicates and coercions between them, originally due to Carlson (1980). Frames represent an additional kind of object which can be an argument to predicates in natural language.

despite Barsalou's interest in *ad hoc* categories. Barsalou (1991), for example, sees *ad hoc* goal-derived categories as important in providing the mapping from frames to world models. Thus while categories are created on the fly, the frames seem less dynamic, even if they are learned over time. Here, however, in talking of coercion we are considering creating frames on the fly. It seems reasonable to say that *some* of the frame types we have available are a permanent part of our general cognitive resources. However, it also seems reasonable to say that *other* frame types can be created *ad hoc* for the purposes at hand and that our ability to do this is exploited in cases of coercion.

## 4 Conclusion

We have discussed a simple-minded theory of the perception of objects and situations couched in terms of a theory of types which takes inspiration from Martin-Löf type theory. As part of this we introduced the notion of record type as corresponding to types of situations like boy-hugging-dog situations where we do not require particular individuals to be involved in the situation. We also suggested that such record types could correspond to types of individuals and raised (but did not solve) issues of individuation which relate to those which have been discussed by Sutton and Filip.

We suggested that such record types can be used to model frame types and that they relate to both the Fillmorean notion of frame and that put forward by Barsalou together with linguistic developments of this notion carried out in Düsseldorf. Despite the fact that the origins of our notion of frame came from Fillmore, the fact that we take a cognitive view of our type theoretic analysis perhaps makes them appropriate for Barsalou's notion.

We discussed work on the Partee puzzle using such frames which seems similar in spirit to Löbner's recent work using frames to analyze the same puzzle. We also pointed out that the techniques we are using seem to have a correspondence to techniques used by Kallmeyer and colleagues, although more detailed investigation would be required to show a general relationship.

Finally, we suggested that the Partee puzzle is not limited to a restricted number of frame level properties but that individual level properties seem to be able to be coerced into frame level properties. This suggests that the frames that we have available as cognitive resources are not necessarily stable but apparently can be created *ad hoc* to meet requirements at hand. This is perhaps an aspect of frames that was discussed neither by Fillmore nor Barsalou.

**Acknowledgements** I would like to thank two anonymous referees for careful and insightful comments. This work was supported by a grant from the Swedish Research Council (VR project 2014-39) for the establishment of the Centre for Linguistic Theory and Studies in Probability (CLASP) at the University of Gothenburg.

## References

- Barsalou, L. W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 27, pp. 1–64). Academic Press.
- Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In A. Lehrer & E. F. Kittay (Eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization* (pp. 21–74). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Barwise, J. (1989). *The situation in logic*. Stanford: CSLI Publications.
- Barwise, J., & Perry, J. (1983). *Situations and attitudes*. Cambridge, Mass: Bradford Books, MIT Press.
- Carlson, G. N. (1980). *Reference to kinds in English*. New York: Garland Publishing.
- Chatzikyriakidis, S., & Cooper, R. (2018). Type theory for natural language semantics. In *Oxford research encyclopedia of linguistics*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.329>.
- Cooper, R. (2010). Frames in formal semantics. In H. Loftsson, E. Rögnvaldsson, & S. Helgadóttir (Eds.), *IceTAL 2010*. Springer.
- Cooper, R. (2012). Type theory and semantics in flux. In R. Kempson, N. Asher, & T. Fernando (Eds.), *Handbook of the philosophy of science, Volume 14: Philosophy of linguistics* (pp. 271–323). Elsevier BV (General editors: Dov M. Gabbay, Paul Thagard and John Woods).
- Cooper, R. (2014). How to do things with types. In V. de Paiva, W. Neuper, P. Quaresma, C. Retoré, L. S. Moss, & J. Saludes (Eds.), *Joint Proceedings of the Second Workshop on Natural Language and Computer Science (NLCS 2014) & 1st International Workshop on Natural Language Services for Reasoners (NLSR 2014) July 17–18, 2014 Vienna, Austria* (pp. 149–158). Center for Informatics and Systems of the University of Coimbra.
- Cooper, R. (2016). Frames as records. In A. Foret, G. Morrill, R. Muskens, R. Osswald, & S. Pogodalla (Eds.), *Formal Grammar: 20th and 21st International Conferences FG 2015, Barcelona, Spain, August 2015, Revised Selected Papers FG 2016, Bozen, Italy, August 2016, Proceedings, number 9804 in LNCS* (pp. 3–18). Springer.
- Cooper, R. (2017a). Adapting type theory with records for natural language semantics. In S. Chatzikyriakidis & Z. Luo (Eds.), *Modern perspectives in type-theoretical semantics, number 98 in Studies in linguistics and philosophy* (pp. 71–94). Springer.
- Cooper, R. (2017b). Neural TTR and possibilities for learning. In S. Dobnik & S. Lappin (Eds.), *CLASP Papers in Computational Linguistics: Proceedings of the Conference on Logic and Machine Learning in Natural Language (LaML 2017)*, Gothenburg, 12–13 June, Volume 1 of *CLASP Papers in Computational Linguistics*, Gothenburg, Sweden. Department of Philosophy, Linguistics and Theory of Science (FLOV), University of Gothenburg, CLASP, Centre for Language and Studies in Probability.
- Cooper, R. (2019). Representing types as neural events. *Journal of Logic, Language and Information*. <https://doi.org/10.1007/s10849-019-09285-4>.
- Cooper, R. (in prep). Type theory and language: From perception to linguistic communication. Draft of book chapters available from <https://sites.google.com/site/typetheorywithrecords/drafts>.
- Cooper, R., & Ginzburg, J. (2015). Type theory with records for natural language semantics. *Lappin and Fox, 2015*, 375–407.
- Fernando, T. (2004). A finite-state approach to events in natural language semantics. *Journal of Logic and Computation*, 14(1), 79–92.
- Fernando, T. (2006). Situations as strings. *Electronic Notes in Theoretical Computer Science*, 165, 23–36.
- Fernando, T. (2008). Finite-state descriptions for temporal semantics. In H. Bunt & R. Muskens (Eds.), *Computing meaning, Volume 3, Volume 83 of Studies in linguistics and philosophy* (pp. 347–368). Springer.
- Fernando, T. (2009). Situations in LTL as strings. *Information and Computation*, 207(10), 980–999.
- Fernando, T. (2011). Constructing situations and time. *Journal of Philosophical Logic*, 40, 371–396.

- Fernando, T. (2015). The semantics of tense and aspect: A finite-state perspective. In Lappin & Fox (Eds.).
- Fillmore, C. J. (1982). Frame semantics. *Linguistics in the morning calm* (pp. 111–137). Seoul: Hanshin Publishing Co.
- Fillmore, C. J. (1985). Frames and the semantics of understanding. *Quaderni di Semantica*, 6(2), 222–254.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton: Mifflin and Company.
- Kallmeyer, L., & Osswald, R. (2013). Syntax-driven semantic frame composition in lexicalized tree adjoining grammars. *Journal of Language Modelling*, 1(2), 267–330.
- Kallmeyer, L., Osswald, R., & Pogodalla, S. (2017). Quantification in frame semantics with binders and nominals of hybrid logic. *Journal of Language Modelling*, 5(2), 357–383.
- Lappin, S., & Fox, C., (Eds.). (2015). *The handbook of contemporary semantic theory* (2nd ed.). Wiley-Blackwell.
- Larsson, S. (2013). Formal semantics for perceptual classification. *Journal of Logic and Computation*, 25(2), 335–369.
- Löbner, S. (2014). Evidence for frames from human language. In T. Gamerschlag, D. Gerland, W. Petersen, & R. Osswald (Eds.), *Frames and concept types*. Studies in linguistics and philosophy (Vol. 94, pp. 23–68). Heidelberg, New York: Springer.
- Löbner, S. (2015). Functional concepts and frames. In T. Gamerschlag, D. Gerland, R. Osswald, & W. Petersen (Eds.), *Meaning, frames, and conceptual representation* (pp. 15–42). Düsseldorf University Press.
- Martin-Löf, P. (1984). *Intuitionistic type theory*. Naples: Bibliopolis.
- Maurin, A.-S. (2016). Tropes. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2016 ed.). Metaphysics Research Lab, Stanford University.
- Moltmann, F. (2013). *Abstract objects and the semantics of natural language*. Oxford University Press.
- Montague, R. (1974). *Formal philosophy: Selected papers of Richard Montague*. New Haven: Yale University Press (ed. and with an introduction by Richmond H. Thomason).
- Nordström, B., Petersson, K., & Smith, J. M. (1990). Programming in Martin-Löf's type theory. In *International series of monographs on computer science* (Vol. 7). Oxford: Clarendon Press.
- Ranta, A. (1994). *Type-theoretical grammar*. Oxford: Clarendon Press.
- Sutton, P. R., & Filip, H. (2017). Individuation, reliability, and the mass/count distinction. *Journal of Language Modelling*, 5(2), 303–356. <https://doi.org/10.15398/jlm.v5i2.144>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# **Conceptualizing Eventualities**

# An XMG Account of Multiplicity of Meaning in Derivation



Marios Andreou and Simon Petitjean

**Abstract** In this paper, we tackle the issue of multiplicity of meaning in derivation using Frame Semantics and eXtensible MetaGrammar (XMG). We use corpus extracted data to identify the range of readings *-al* derivatives exhibit and identify prominent constraints on the types of situations and entities *-al* targets. These constraints have the form of type constraints and specify which arguments in the frame of the verbal base are compatible with the referential arguments of the derivative. The introduction of these constraints into the semantics of an affix allows one to predict and generate those readings which are possible for a given derivative and, at the same time, rule out those readings which are not possible. Finally, as a proof of concept, we model these constraints using XMG, and check whether the output resulting of this XMG description is consistent with the range of readings observed in the corpus.

**Keywords** Derivation · Polysemy · Constraints · Frame semantics · Extensible metagrammar

## 1 Introduction

More often than not, the products of derivational processes are interpreted in more than one way. This multiplicity of meaning is particularly evident in deverbal nominalizations (Lieber 2004; Lieber and Andreou 2018; Rainer 2014; Andreou and Petitjean 2017; Plag et al. 2018). Derived words that are based on the suffix *-al*, for example, may denote either situations (e.g. *removal* “the act of removing”) or entities (e.g. *rental* “the thing one rents”).

In this paper, we focus on deverbal nominalizations with the suffix *-al* that are based on causation events. Causation events have a rich bipartite structure which

---

M. Andreou (✉) · S. Petitjean  
Institute of Linguistics and Information Science, Heinrich-Heine-Universität, 40204 Düsseldorf,  
Germany  
e-mail: [Marios.Andreou@uni-duesseldorf.de](mailto:Marios.Andreou@uni-duesseldorf.de)

<i>ball</i>	
SHAPE	<i>round</i>

**Fig. 1** Partial frame for *ball*

captures complex relationships between situations (events and states) and entities. This complex structure allows one to identify and test constraints that might affect the types of arguments which *-al* targets.

The aim of the paper is threefold. First, to best describe the behavior of *-al* on causation events and, thus, capture the multiplicity of meaning exhibited by *-al* nominalizations. Second, to identify prominent constraints on the types of situations and entities *-al* targets. This will allow us to inform the discussion on the way one can greatly reduce overgeneration of readings. In particular, the identification of constraints will be a contribution to the literature on the way one can predict and generate those readings which are possible for a given derivative and, at the same time, rule out those readings which are not possible (Lieber 2004; Booij 2010; Rainer 2014; Andreou and Petitjean 2017; Plag et al. 2018). Third, to best model these constraints using XMG.

Our approach is based on the framework of Frame Semantics as developed in Petersen (2007), Kallmeyer and Osswald (2013), and Löbner (2013, 2014, 2015).<sup>1</sup> A frame is a general format of mental representations of concepts which is also applicable to linguistic phenomena. It is a recursive attribute-value structure that provides information about the referent of the frame. Attributes are applied to a given possessor in a frame structure and assign a value to it.<sup>2</sup> To provide an example, Fig. 1 gives the partial frame for *ball* in the form of an attribute-value matrix.

The referent of the frame in Fig. 1 is *ball*. The attribute-value matrix illustrates that *ball* has an attribute SHAPE and that this attribute assigns the value *round* to the referent of the frame. Thus, the shape of the referent of the frame, i.e. *ball*, is round.

Word formation in Frame Semantics is generally treated in terms of referential shifts (Löbner 2013; Plag et al. 2018). In particular, reference is shifted from the original referent to a new referent. For example, as we will see in the analysis, the suffix *-al* can target particular arguments of the base verb and shift reference from the original referent (i.e. causation event) to a new referent (e.g. theme). As recently shown by a number of studies on nominalizations (Lieber 2004, 2016; Kawaletz and Plag 2015; Andreou and Petitjean 2017; Plag et al. 2018), not all arguments of the verb can be targeted by affixation. The identification of prominent constraints on

<sup>1</sup> Frames also figure in works on Lexical Functional Grammar (Bresnan 2001), Head-Driven Phrase Structure Grammar (Pollard and Sag 1994), and Sign-based construction grammar Sag (2012). Fillmore's frames (Fillmore 1982) are used in the FrameNet project (Fillmore and Baker 2010). In the present paper, we will use Frames as defined in the work of Petersen (2007), Kallmeyer and Osswald (2013), and Löbner (2013, 2014, 2015), which is inspired by the work of Barsalou (1992a); Barsalou, (1992b); Barsalou (1999).

<sup>2</sup> Attributes will be given in small capitals and values in italics.

the types of arguments that can be targeted by a particular affix is still an open issue and has implications for the way we describe, model, and implement a particular derivational process in XMG.

What is XMG? XMG (eXtensible MetaGrammar, Crabbé et al. (2013)) is a modular and extensible tool used to generate various types of linguistic resources from an abstract and compact description. This description, the metagrammar, relies on the concepts of logic programming and constraints. XMG comes with a system of dimensions, allowing one to separate the different levels of linguistic description (e.g. syntax and semantics), and providing dedicated languages adapted to the structures the user wishes to generate. In this work, the dimension we used is the <frame> dimension, proposed in Lichte and Petitjean (2015), where semantic frames can be described using typed feature structure descriptions.

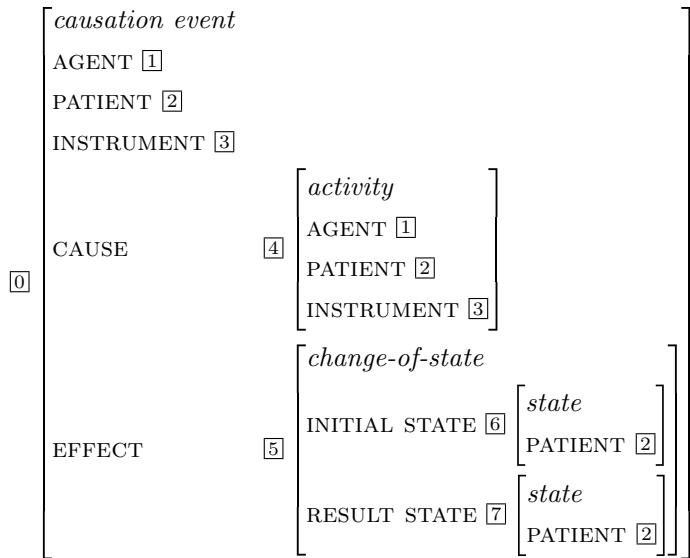
The rest of this paper is structured as follows: In Sect. 2, we describe and analyze the behavior of *-al* nominalizations in context. This will allow us to identify prominent constraints on the types of situations and entities that can be targeted by *-al*. In Sect. 3, we provide an analysis of the multiplicity of meaning exhibited by *-al* nominalizations in XMG. Section 4 concludes the paper.

## 2 Data and Analysis

In this paper, we follow the classification of VerbNet (Kipper-Schuler 2006) that is inspired by the classification of Levin (1993) and we focus on the suffix *-al* on causation events. In particular, we examine the following verb classes: put verbs (e.g. *bury*), remove verbs (e.g. *remove*), banish verbs (e.g. *recuse*), deprive verbs (e.g. *deprive*), send verbs (e.g. *transmit*), contribute verbs (e.g. *betroth*), verbs of future having (e.g. *bequeath*), equip verbs (e.g. *redress*), get verbs (e.g. *procure*), obtain verbs (e.g. *retrieve*), amuse verbs (e.g. *arouse*), verbs of change of state (e.g. *disperse*), free verbs (e.g. *acquit*), addict verbs (e.g. *dispose*), and base verbs (e.g. *construe*).

We chose to work with causation events since these verbs have a rich bipartite structure which captures complex relationships between situations and entities. Thus, by using causation events as a testbed we can identify constraints on the types of situations and entities *-al* targets. In particular, we can ask the following question: Are all situations and entities able to be targeted by *-al* affixation or are there general constraints on the types of arguments *-al* targets?

A typical causation event comes with a bipartite structure that comprises a CAUSE and an EFFECT (Kallmeyer and Osswald 2012; Plag et al. 2018). It involves a relationship between situations and entities in which a particular entity (e.g. an originator in the sense of (Borer 2014)) causes another entity (i.e. a theme) to go from an initial situation to a result situation (Lieber 2004; Levin 1993; Rappaport Hovav and Levin 2008). The following two attribute-value matrices illustrate this state of affairs. Figure 2 gives the structure of a change of state verb such as *renew* and Fig. 3 illustrates the structure of a verb of change of possession such as *bequeath*.



**Fig. 2** Change of state verbs

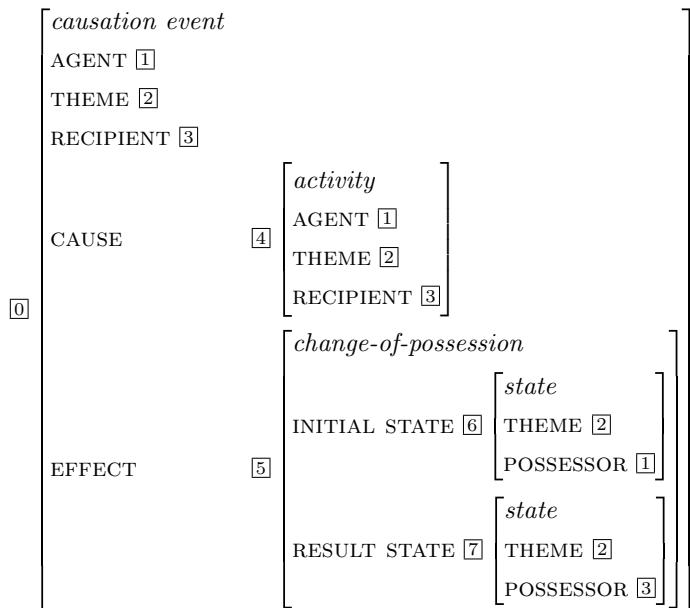
Figure 2 models that *renew* comes with a bipartite structure that comprises a CAUSE (i.e. *activity*) and an EFFECT (i.e. *change-of-state*). In particular, *renew* involves a relationship between the participants agent, patient, and instrument, in which the agent causes the patient to go from an initial state to a result state.

Another example which shows that causation events generally involve two sub-events, a cause and an effect, is given in Fig. 3 which models a future having verb such as *bequeath*. This verb describes caused possession of the kind ‘x causes y to have z’, in which x is the agent, y is the recipient, and z is the theme (Goldberg 1995; Jackendoff 1990; Rappaport Hovav and Levin 2008). Thus, Fig. 3 models this state of affairs as a relationship between an agent, a theme, and a recipient, in which there is an initial situation in which the agent has possession of the theme, and a result situation in which the recipient has possession of the theme (Andreou and Petitjean 2017).

Let us now present the findings of our study with respect to possible readings of -al nominalizations. We use data from the Corpus of Contemporary American English (COCA, (Davies 2008)). Among the readings we find in causation events, the most productive are the event and result readings. (1) includes event readings and (2) provides result readings.

### (1) Event reading

- a. One can perhaps gain a further glimpse of this sort of process of **construal** in a 1979 conversation of Serra, Annette Michelson, and Clara Weyergraf. Michelson began the interview by asking Serra how and when he came to filmmaking. (COCA ACAD 2015)



**Fig. 3** Verbs of change of possession

- b. This results in delays in the **disbursal** and utilization of funds—especially at the Gram Panchayat level. (COCA ACAD 1998)
  - c. If it is morally unacceptable to repatriate even a convicted illegal alien criminal, then it is all the more unacceptable to repatriate someone who “merely” has crossed the border illegally. This undermining of alien **removals** is behind the constant protests demanding to “stop deportations now.” (COCA MAG)
- (2) Result reading
- a. Introverts proved more able to focus on the task of color identification while disregarding the emotional content and had significantly better reaction times. Concludes Haas: Introverts, who exhibit a higher resting state of **arousal**, “don’t need the same kind of outside entertainment.” (COCA MAG 2010)
  - b. At the same time as it emerged that Fitzroy was terminally ill with ‘a rapid consumption’, Henry learned of Margaret Douglas’s **betrothal** to Thomas Howard. (COCA MAG 2013)
  - c. Smith, 54, is the nephew of a slain American president. As a younger man, he was the defendant in a salacious Palm Beach rape trial that ended in his **acquittal**, though not before the nation devoured stories of late-night, alcohol-fueled carousing that included then-Sen. (COCA NEWS 2014)

In the examples in (1), the nominalization lexicalizes the event denoted by the verb. This type of nominalization is also referred to as ‘transpositional’ in that the nominalization ‘transposes’ (recategorizes) the word from verb to noun without altering the sense of the verbal base. Thus, *construal*, *disbursal*, and *removals* can be paraphrased as “event/process of construing”, “event/process of disbursing”, and “event/process of removing”, respectively.

In the examples in (2) the nominalization has a result reading<sup>3</sup> in that it lexicalizes “the outcome of verb-ing”. Thus, *arousal*, *betrothal*, and *acquittal* can be paraphrased as “the (result) state of arousing”, “the outcome of betrothing”, and “the outcome of acquitting”.

Observe that in both (1) and (2), contextual cues may guide us to a particular reading. For example, *the process of construal* flags a transpositional eventive reading and *a higher state of arousal* guides us towards a result state reading.

One may also find *-al* nominalizations that lexicalize the inanimate theme, that is, “the thing verb-ed, the thing affected by verb-ing”. Consider the examples in (3).

### (3) Inanimate theme

- a. Planning for and pursuing invoices is necessary in any case. After **renewals** are paid in July or August (or the first two months), September (or the third month) is a good time to start setting up projection reviews for these resources. (COCA ACAD 2015)
- b. The room was technically full of locals, people from Bianca’s life before she headed West, friends who crossed the bridge searching for more affordable **rentals** in Williamsburg or Long Island City. (COCA FIC 2015)
- c. In any case, your best bet is to roll the money into a traditional IRA; otherwise, you’ll get a big tax bill. Smaller **withdrawals** from the IRA, on the other hand, will likely be taxed at a lower rate. (COCA MAG)

In (3), we observe that *renewals* are “the things one renews (e.g. subscriptions)”, *rentals* are “the things that someone rents (e.g. a house, an apartment)”, and *withdrawals* are “the things one withdraws (i.e. money)”.

A closer inspection of the data in (1)–(3) reveals that the suffix *-al* can manipulate the frame of a verb and target certain arguments of it. In particular, it can target the causation event argument, the result situation argument, and the theme argument. Thus, the referent of a form derived by *-al* can be identified with some of the arguments of the verbal base, but not all of them. Observe, for instance, that the referent of *-al* derivatives is never the agent, the recipient, the cause, the effect or the initial situation.

In what follows, we undertake the nontrivial task of identifying possible constraints on the types of entities and situations *-al* targets.

As far as entities are concerned, there seems to be a constraint on the animacy of the referent of *-al* nominalizations. In particular, the referent of *-al* nominalizations

---

<sup>3</sup>The examples b. and c. are bounded, in that they happened in the past. For more on aspect in nominalizations the interested reader is referred to Lieber and Andreou (2018).

cannot be [+animate]. This explains why we find inanimate theme readings but not agentive readings.

In what follows, we test this constraint on animacy. Consider the following examples:

(4) a. Agentive reading

The path down to the sea is shaded by lemon groves. There is also an elevator to the private beach, where a saltwater pool, sun decks, a bar and seaside restaurant, along with a well-equipped gym and **boat rentals**, await. (COCA MAG 2001)

b. Instrument reading

If I hadn't read the article in your magazine, my precious dogs would be in continued danger. Enclosed is my **renewal**. Thanks for the great information. (COCA MAG 2003)

c. Asset reading

The farmer who owned the barn had asked - and received - a thousand dollars in **rental**. (COCA FIG 2004)

Although the examples in (3) are not primary readings of *-al* nominalizations, they can, nevertheless, inform the discussion on the constraint on animacy. In (4-a), *boat rentals* has an agentive reading. This seems to militate against the hypothesis that the referent of *-al* nominalizations cannot be [+animate]. On closer inspection, however, the context suggests that the referent of *boat rentals* is inanimate. It is the company that rents boats. In any case, this reading is highly lexicalized. In (4-b), *renewal* is interpreted as an instrument since it is the participant in the *renew* event that is manipulated by the agent, and with which an intentional act is performed. In our example, it is the form of renewal of subscription. Thus, the referent of *renewal* is inanimate. Finally, the argument that seems to be lexicalized in (4-c) is the asset argument, that is the value of something. In our example, *rental* lexicalizes this argument since its reading can be paraphrased as "the amount of money one has to pay for renting the barn". To sum up, the examination of secondary readings of *-al* nominalizations confirms the hypothesis that there is a constraint on animacy on the referent of *-al* forms.

Let us now turn to situations. Is there a constraint on the types of situations that can be targeted by *-al*? As mentioned above, the structure of causation events typically includes the causation event argument, a cause, an effect, an initial situation, and a result situation. In our data, there are no cases in which the cause, the effect or the initial situation are targeted by *-al*. As shown in (1) and (2), *-al* nominalizations in our data give rise only to transpositional eventive readings and result situation readings. Let us elaborate upon the latter reading, i.e. result situation. The result situations described by the various subclasses in our data are not homogeneous. In particular, verbs such as *arouse* describe a change of emotional state, verbs such as *bequeath* describe a change of possession, and verbs such as *remove* describe a change of location. Are all these situations able to be targeted by *-al*?

Our data suggest that the only result situation that is compatible with *-al* is the result state. The only example in which we identified a different reading is given below:

- (5) In a **burial** in Gyeongju, South Korea, archaeologists uncovered armor of a fifth-century A.D. warrior and his horse, as well as dozens of serving vessels used in traditional burial rituals. (COCA ACAD 2009)

This reading involves the put verb *bury* which describes a change of location. The use of *burial* with the reading of result location (e.g. tomb, grave), however, is highly lexicalized and only used in archeology. Thus we can safely conclude that the referential argument of *-al* forms is not compatible with arguments of the type location.

The identification of these constraints allows one to comment on the way one can handle multiplicity of meaning in derivation. In the relevant literature (Lieber 2004; Booij 2010; Rainer 2014; Andreou and Petitjean 2017; Plag et al. 2018), there are two approaches to multiplicity of meaning in derivation. Under the first approach, i.e. monosemy, more concrete meanings of affixes derive from a general highly underspecified meaning that is capable of taking into account all possible readings of an affix.

Applying the monosemy approach to *-al* consists in reducing the multiplicity of meaning by identifying meanings that are shared by all *-al* derivatives. As follows from the analysis of our data, *-al* derivatives denote (a) eventualities (e.g. event ‘transpositional’ readings), and (b) entities (e.g. inanimate theme readings). Thus, the abstract core meaning of *-al* can be characterized as ‘eventuality or entity having to do with X’ (with ‘X’ denoting the base).

Monosemy approaches to the semantics of derivation are confronted with two problems. The first problem is that it is very hard to establish a unitary meaning for an affix. In particular, the aim of monosemy approaches is to reduce multiplicity of meaning by postulating a unitary abstract meaning. Forms derived by *-al*, however, denote both eventualities and entities. Thus, the disjunction ‘eventuality or entity’ that is needed in order to capture the multiplicity of meaning of *-al* derivatives reveals that the desirable underspecified meaning of affixes cannot always be reduced to a single unitary meaning.

The second problem with the monosemy approach to the semantics of derivation is (massive) overgeneration. As we saw earlier, the abstract meaning for *-al* informs us that *-al* forms denote both eventualities and entities. What kind of predictions follow from the abstract meaning ‘eventuality or entity having to do with X’? This particular formulation of the abstract meaning of *-al* leads one to expect that *-al* derivatives could in principle denote all entities and all eventualities. Our data, however, suggests that not all entities and not all eventualities can be denoted by *-al* derivatives. For instance, the referent of an *-al* derivative may be the inanimate theme (e.g. *money* in the case of *withdrawal*) but not the agent.

Under the second approach, i.e. polysemy, there is multiplicity of meaning in word formation patterns. Given the architecture of Frame Semantics, the multiplicity

of readings exhibited by *-al* nominalizations can be captured with the use of an inheritance hierarchy of lexeme formation rules (Riehemann 1998; Koenig 1999; Booij 2010; Bonami and Crysmann 2016; Plag et al. 2018). Inheritance hierarchies allow one to generalize over derived formations and capture shared characteristics between them as we show in Fig. 4.

Figure 4 gives the inheritance hierarchy of lexeme formation rules ('lfr') for deverbal nominalizations ('v-n') in *-al*. This hierarchy involves two dimensions, namely phonology (PHON) and semantics (SEM). The first dimension, i.e. phonology, is shared by all *-al* nominalizations. In particular, all *-al* nominalizations have the phonology /1+al/. Boxed numerals such as [1] are called tags and are used in feature structures to indicate structure sharing, that is, to show that the respective values are identical. In Fig. 4, this means that the value of the first part of the phonology of the derived lexeme is identical to the value for the phonology of the base. The second part of the phonology of the derived lexeme is, of course, contributed by the affix, i.e. /al/.

Although *-al* nominalizations are based on the same phonological pattern, their semantics differs. The semantic dimension in the inheritance hierarchy in Fig. 4 captures the different readings exhibited by *-al* forms. In accordance with the analysis suggested by our data, when the reference of a form in *-al* is identified with the event argument ('evt') of the base, we get an eventive 'transpositional' reading and when it is identified with the result state argument ('r-st') of the base, we get a result state reading. In a similar vein, a theme reading arises when the reference of an *-al* nominal is identified with the theme argument ('thm') of the base, an instrument reading when it is identified with the instrument argument ('inst') of the base, and finally an asset reading when it is identified with the asset argument ('ast') of the base. The lowest level of Fig. 4 shows that *-al* forms inherit their characteristics from both dimensions, i.e. phonology and semantics. In particular, all *-al* forms share the same phonology, but their semantics differs.

In this section, we identified the range of readings available to *-al* forms and described the way this range could be accounted for under the monosemy and polysemy approach. In the next section, we will use the type constraints we identified in this section, in order to predict and generate those readings which are possible for an *-al* form and, at the same time, rule out those readings which are not possible.

### 3 XMG Implementation

The XMG compiler is a tool which has already been used to generate a wide range of linguistic resources, focusing on different levels of linguistic description, such as syntax and semantics, or even interfaces between them. Syntactic resources developed with XMG are tree-based grammars such as Tree Adjoining Grammars (Crabbé 2005; Kallmeyer et al. 2008; Gardent 2008 for instance) or Interaction Grammars such as Perrier (2007). Other types of resources include lexicons of

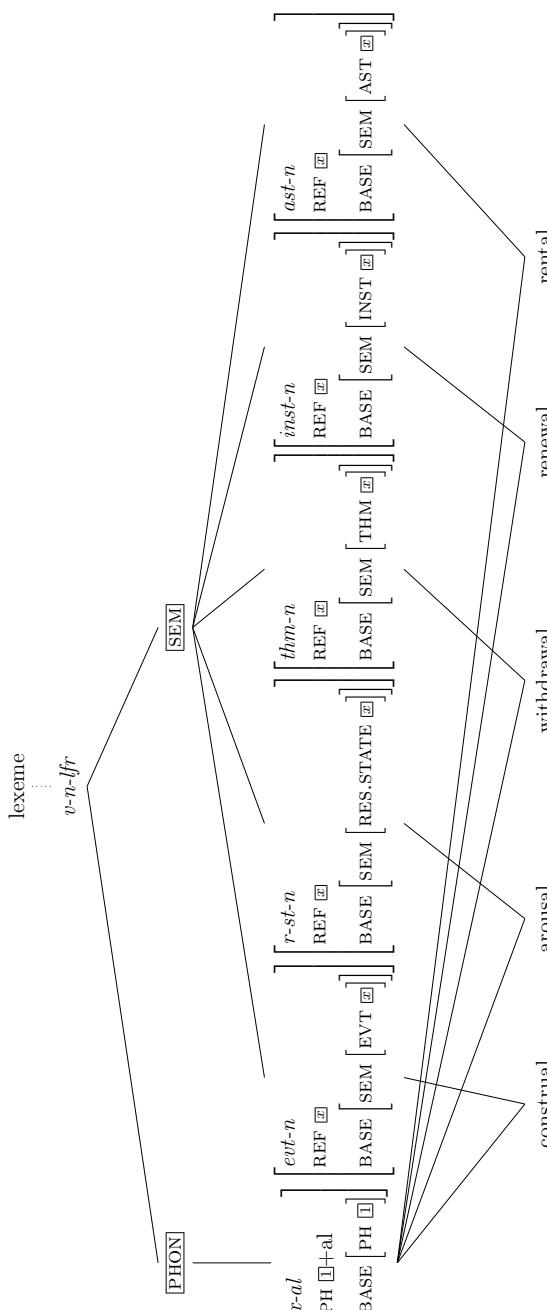


Fig. 4 Inheritance hierarchy of lexeme formation rules for the suffix *-al*

fully inflected forms, which were generated from morphological descriptions as in Duchier et al. (2012), or frame-based semantic descriptions. In this work, even though we are interested in both morphology and semantics, we will only focus on the description of the semantics. On the morphological side, the description is trivial as it only consists in combining a verb and a given affix.

An XMG implementation is a program (called metagrammar) composed of a set of classes, which are reusable abstractions. A class describes a partial linguistic structure, which is in our case the frame for a given class of verbs. Classes can be reused by other classes (imported), to add information to the partial description. This is what will be done by the classes modeling derivations: they will import the descriptions of the verb frames and augment them by defining the semantic reference corresponding to one reading of the derivation. The descriptions shown in this article mainly consist of typed feature structures. By using unification variables in their description, the feature structures are combined to describe more complex frames. An XMG program is non-deterministic: it uses underspecification and disjunction, meaning that every class can describe zero, one or more structures. When the metagrammar is processed by the XMG compiler, all the structures described in the classes are computed and written into an output file (using the XML or JSON format).

The implementation that we present aims at generating the frames corresponding to all the attested readings for the derivations. For space limitations, below we focus on two classes, namely, verbs of change of possession and verbs of change of state. The proposed analysis can, nevertheless, be extended to additional verb classes in a similar and straightforward manner.

We first need to describe the frame given in Fig. 3, by means of a XMG class which we will name *rent*. This abstraction describes the class of verbs of change of possession:

```

class rent
export ?X0
declare ?X0 ?X1 ?X2 ?X3 ?X4 ?X5 ?X6 ?X7
{<frame>{
    ?X0[causation,
        agent: ?X1[entity, animacy:[animate]],
        theme: ?X2,
        recipient: ?X3[entity],
        cause: ?X4[activity,
            agent: ?X1,
            theme: ?X2,
            recipient: ?X3[entity, animacy:[animate]]]
        ],
    effect: ?X5[change_of_possession,
        initial-state: ?X6[initial_state,
            theme: ?X2[entity],
            possessor: ?X1],
        result-state: ?X7[result_state,
            ]
    ]
}

```

```

    theme:?X2[entity],
    possessor:?X3 ]]
}
}
```

where the first lines define the set of unification variables which can be used within the class (**declare**) and outside of it (**export**). These variables can be matched with any value or structure described in the metagrammar (a feature structure, the value for a specific attribute, a syntactic node, etc). <**frame**> means that the description belongs to the Frame Semantics dimension. The structure described in the frame dimension, labeled by ?X0, is a straightforward translation of the one in Fig. 3, with the addition of information on animacy, where all variables ?X0,...,?X7 stand for the boxed numbers from [0] to [7]. The only variable which can be accessed outside of the class is ?X0 (cf. **export** ?X0). In the same fashion, we define the class of verbs of change of state shown in Fig. 2.

```

class renew
export ?X0
declare ?X0 ?X1 ?X2 ?X3 ?X4 ?X5 ?X6 ?X7
{<frame>{
    ?X0[causation,
        agent: ?X1[entity, animacy:[animate]],
        patient: ?X2,
        instrument: ?X3[entity],
        cause: ?X4[activity,
            agent:?X1,
            patient:?X2,
            instrument:?X3[entity, animacy:[animate]]
        ],
        effect: ?X5[change_of_state,
            initial-state: ?X6[initial_state, patient:?X1],
            result-state: ?X7[result_state, patient:?X3]]]
    }
}
```

To define the scope-over relation mentioned earlier, we can use a new abstraction (a class we will name *al\_nominal*). This class, as its name suggests, models the semantics of *-al* derivatives, which for the purposes of this first example are based on verbs of change of possession.

```

class al_nominal
import rent[]
declare ?Ref
{
    <frame>{
        [al-lexeme,
```

```

m—base:[event,
      sem:?X0]
ref:?Ref
]
;
?X0 >* ?Ref;
}
}

```

With **import rent[]** we make the structure defined in the class rent available in the current class, together with its variables (we can refer only to the foreign variable ?X0 in the current class as only this variable is exported by rent). The operator  $>$  is used to specify an additional constraint on the frame: the left operand is a frame and the right operand must be one of the values of its attributes. Here, we use the reflexive transitive closure of this operator,  $>*$ , which means that there must be a path (as it would be in a graph representation<sup>4</sup> of the frame) from the root ?X0 to the semantic reference ?Ref. Concretely, the compiler will try to generate structures where the reference is identified with another label, starting with the whole frame (?X0), and then exploring all of its subparts, recursively. This is comparable to functional uncertainty in LFG as defined by Kaplan and Maxwell (1988), even though we believe it to be more general: when using only the operator  $>*$ , the reference will be able to unify with every possible subpart, totally independently from the attributes composing the path. As in the solution proposed by Krieger et al. (1993) to implement functional uncertainty, type constraints are essential: they will be the main way for us to control which subparts can be identified with the semantic reference.

As said previously, with this description, all possible subparts of the feature structures are possible candidates to be identified with the reference, and as a consequence, readings such as initial state (which should be ruled out) are also generated when this first version of the metagrammar is executed.

In this first implementation we modeled an approach to multiplicity of meaning which is close to a version of the monosemy approach under which there are no constraints on types, and showed that it leads to massive overgeneration. In the next section we focus on the second approach to multiplicity of meaning: polysemy.

An open question is how we can model the polysemy approach in XMG and constrain possible readings. We suggest that there are two ways to tackle this issue. First, via a fully specified (and explicit) rule, which will replace the scope over relation in the previous class al\_nominal:

```
{?X0=?Ref | ?X2=?Ref | ?X7=?Ref}
```

---

<sup>4</sup>An attribute-value matrix can be seen as a directed graph in which every attribute-value pair is an edge labeled by the attribute and pointing to the node representing the value.

where  $|$  and  $=$  are respectively the disjunction and the unification operators,  $?X0$ ,  $?X2$  and  $?X7$  respectively correspond to the boxed numbers 0, 2 and 7 of Fig. 3, and  $?Ref$  is a variable representing the semantic reference.

Under this approach, possible readings are considered as generalizations over already attested derivatives. Thus, agent, recipient, and initial state readings are ruled out since they are not part of the possible readings in the fully-specified-rule; the rule models readings that are already attested in *-al* derivatives. However, this implementation is totally specific to a given class of verbs, here the one described in the class *rent*. More XMG code would have to be written for the derivation of other verb classes, where the reference would be identified with different unification variables. In our case, we used consistent variable namings in the class *renew* (the variables corresponding to the attested readings are also  $?X0$ ,  $?X2$  and  $?X7$ ), making it easily compatible with this implementation, but it would not be as straightforward for frames with different numbers of features. For example, for a verb class where the *-al* nominalization has four different readings, a different XMG class with four alternatives of variable unifications would have to be used.

Another way to model the polysemy approach in XMG is the introduction of an underspecified rule with constraints on types. Only the types of the feature structures will determine if one reading should be valid or not, which means that we do not need to provide explicitly the set of variables that may be unified with the semantic reference. In the case of our verb classes, the referent of an *-al* nominal can have three possible types: causation, result state, or entity.

```
?X0 >* ?Ref;
{ ?Ref[result_state] | ?Ref[causation]
| ?Ref[entity, animacy:[inanimate]] }
```

Here, the first line is once again the scope over relation, but of course, in this case, only the structures where no type constraint is violated will eventually be generated.

In the second line, we express the fact that the referent of an *-al* derivative can have any of the three types previously stated. In the case of an entity, only the theme should be a possible referent. We, therefore, add information about animacy (here, inanimate), which makes the reference of *-al* derivatives incompatible with frames of type animate, such as the agent and the recipient. This is in accordance with findings in the literature on possible constraints on animacy (see Kawaletz and Plag (2015) on the suffix *-ment*). When the referent of an *-al* derivative is a state, the type *result\_state* is given to prevent unification with the initial state frame (of type *initial\_state*). This way, agent, recipient, and initial state readings are ruled out because frame unification only succeeds if types are compatible. The type constraints (for example incompatibility of event and entity) are also specified in the metagrammar. This is done globally, meaning that the type constraints will apply to all the structures described in the metagrammar. The constraints defining our type hierarchy are introduced by the keyword **frame**—constraints as follows:

```

frame-constraints ={
    event -> eventuality,
    state -> eventuality,
    state event -> -,
    eventuality entity -> -,
    derived-lexeme -> lexeme,
    ment-lexeme -> derived-lexeme,
    lexeme eventuality -> -,
    eventuality entity -> -,
    causation -> event,
    activity -> event,
    change_of_possession -> event,
    change_of_state -> event,
    causation activity -> -,
    causation change_of_possession -> -,
    causation change_of_state -> -,
    change_of_state change_of_possession -> -,
    experiencer -> entity,
    stimulus -> entity,
    experiencer stimulus -> -,
    initial_state result_state -> -,
    initial_state -> state,
    result_state -> state,
    animate inanimate -> -,
    animate -> animacy,
    inanimate -> animacy,
    animacy eventuality -> -,
    animacy entity -> -,
    entity -> animacy:animacy,
    animacy lexeme -> -
}

```

Three types of constraints are used here, all using the  $\rightarrow$  operator, which can be read as an implication. Subsumption constraints, such as causation  $\rightarrow$  event, mean that an atomic type (here causation) is a subtype of another type (event). The effect of this constraint is that a frame cannot have the type causation without having the type event as well. An incompatibility constraint, such as causation activity  $\rightarrow$   $-$  means that a structure cannot have both of the two given types: here, a frame cannot be of type causation and of type activity. Finally, feature constraints, such as entity  $\rightarrow$  animacy:animacy ensure that all the structures having a given type have a given feature. In our case, structures of type entity will all have an attribute animacy of type animacy. The set of type constraints defines the type signature of the metagrammar.

This implementation is directly compatible with the verbs described in the class renew, and does not depend on the naming of the variables used in the frame

description. Therefore, an XMG abstraction describing verbs from another class, even if it is written by another linguist who uses different naming conventions, can be combined with the al\_nominal class. Of course, for verb classes in which readings are not limited to the same types (causation, result\_state and inanimate entity), new XMG abstractions for *-al* nominalization would have to be written. In these new XMG classes, only the type constraints would differ, and they could be directly reused for all other verb classes with similar behaviors.

## 4 Conclusion

In the present paper, we tackled the issue of multiplicity of meaning in derivation by offering a detailed analysis of *-al* derivatives. We used corpus extracted data to identify the range of readings available to *-al* derivatives and to establish possible constraints on the types of arguments *-al* targets. Finally, we modeled these constraints using XMG.

In a nutshell, we showed that the referent of an *-al* derivative can be identified with certain types of situations and entities, but not all of them. This has implications for the way we model multiplicity of meaning in derivation, since it shows that it is not always possible to reduce the meaning of a particular affix to a single unitary meaning.

Our XMG implementation corroborates the idea that the introduction of constraints into the semantics of an affix allows one to predict and generate those readings which are possible for a given derivative and rule out other readings which are not possible. These constraints have the form of type constraints and specify which arguments in the frame of the verbal base are compatible with the referential argument of the derivative. The introduction of type constraints rules out certain readings because frame unification only succeeds if types are compatible.

In the present paper, we focused on *-al* derivatives. The next step is to apply the proposed analysis to the modeling of other affixes as well. This will allow us to identify which constraints are specific to particular classes or affixes, and which constraints are shared across classes or affixes. For example, the suffixes *-ance*, *-ment*, and *-ure* show similar characteristics to the suffix *-al*, in that the referent of forms derived by these affixes is never [+animate]. They differ, however, from one another with respect to other characteristics. For example, *-ance*, *-ment*, and *-ure* are compatible with the location argument of the verbal base, whereas *-al* is not, and *-ure* is not compatible with the instrument argument of the verbal base, whereas *-ance*, *-ment*, and *-al* are. The main advantages of the metagrammatical framework will become more obvious as the linguistic resource grows: for example, inheritance will help sharing information across classes with similar behaviors.

**Acknowledgements** We would like to thank Ingo Plag, Lea Kawaletz, Curt Anderson, two anonymous reviewers, and the audience at COST (Cognitive Structures: Linguistic, Philosophical and Psychological Perspectives, September 15-17, 2016, Heinrich-Heine-Universität Düsseldorf) for

their comments and suggestions. We gratefully acknowledge that this research has been funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, Collaborative Research Centre 991, Project C08 ‘The semantics of derivational morphology: A frame-based approach’) and by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Project 724411, TreeGraSP ‘Tree rewriting grammars and the syntax-semantics interface: From grammar development to semantic parsing’).

## References

- Andreou M., & Petitjean S. (2017). Describing derivational polysemy with xmg. In Eshkol, I. & Antoine, J.-Y. (Eds.), *Actes de TALN 2017, 24e Conférence sur le traitement automatique des langues naturelles*, Vol. 2, pp. 94–101.
- Barsalou, L. W. (1992b). Frames, concepts, and conceptual fields. In Lehrer, A. & Kittay, E. F. (Eds.), *Frames, fields and contrasts: New essays in semantic and lexical organization*, pp. 21–74. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Barsalou, L. W. (1992a). *Cognitive psychology: An overview for cognitive sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences*, 22, 577–660.
- Bonami, O., & Crysmann, B. (2016). Morphology in constraint-based lexical approaches to grammar. In A. Hippisley & G. T. Stump (Eds.), *Cambridge handbook of morphology* (pp. 609–656). Cambridge: Cambridge University Press.
- Booij, G. (2010). *Construction morphology*. Oxford: Oxford University Press.
- Borer, H. (2014). *Taking form structuring sense, Vol. III*. Oxford: Oxford University Press.
- Bresnan, J. (2001). *Lexical functional syntax*. Malden - Oxford: Blackwell.
- Crabbé, B. (2005). *Représentation informatique de grammaires fortement lexicalisées : Application à la grammaire d’arbres adjoints*. PhD thesis, Université Nancy 2.
- Crabbé, B., Duchier, D., Gardent, C., Le Roux, J., & Parmentier, Y. (2013). *XMG : eXtensible MetaGrammar Computational Linguistics*, 39(3), 1–66.
- Davies, M. (2008). The Corpus of Contemporary American English: 400+ million words, 1990–present.
- Duchier, D., Ekoukou, B. M., Parmentier, Y., Petitjean, S. & Schang, E. (2012). Describing morphologically-rich languages using metagrammars: A look at verbs in Ikota. In: *Workshop on "Language Technology for Normalisation of Less-Resourced Languages", 8th SALTMIL Workshop on Minority Languages and the 4th workshop on African Language Technology*, p. 97–106, Istanbul, Turkey.
- Fillmore, C. J. (1982). Frame semantics. In The linguistic society of Korea. (Ed.), *Linguistics in the morning calm*, pp. 111–137. Seoul: Hanshin Publishing Co.
- Fillmore, C. J., & Baker, C. (2010). A frames approach to semantic analysis. In Heine, B. & Narrog, H. (Eds.), *The oxford handbook of linguistic analysis* pp. 313–340. Oxford: Oxford University Press.
- Gardent, C. (2008). Integrating a unification-based semantics in a large scale lexicalised tree adjoining grammar for French. In: *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pp. 249–256, Manchester, UK: Coling 2008 Organizing Committee.
- Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.
- Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.
- Kallmeyer, L. & Osswald R. (2012). A frame-based semantics of the dative alternation in Lexicalized Tree Adjoining Grammars. In Piñón, C. (Ed.), *Empirical issues in syntax and semantics*, Vol. 9, pp. 167–184. ISSN 1769-7158.

- Kallmeyer, L., & Osswald, R. (2013). Syntax-driven semantic frame composition in lexicalized tree adjoining grammars. *Journal of Language Modelling*, 1(2), 267–330.
- Kallmeyer, L., Lichte, T., Maier, W., Parmentier, Y., & Deller, J. (2008). Developing a TT-MCTAG for German with an RCG-based Parser. In *The sixth international conference on language resources and evaluation (LREC 08)*, pp. 782–789. Morocco: Marrakech.
- Kaplan, R. M., & Maxwell, J. T. (1988). An algorithm for functional uncertainty. In: *Proceedings of the 12th Conference on Computational Linguistics*, Vol. 1, COLING '88, pp. 297–302. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Kawaletz, L., & Plag, I. (2015). Predicting the semantics of english nominalizations: A frame-based analysis of -ment suffixation. In L. Bauer, L. Körtvélyessy, & P. Štekauer (Eds.), *Semantics of complex words* (pp. 289–319). Cham: Springer International Publishing.
- Koenig, J. (1999). *Lexical Relations*. Stanford Monographs in Linguistics. Cambridge University Press.
- Krieger, H.-U., Pirker, H., & Nerbonne, J. (1993). Feature-based allomorphy. In: *Proceedings of the 31st Annual Meeting on Association for Computational Linguistics*, ACL '93, pp. 140–147. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago, IL: University of Chicago Press.
- Lichte, T., & Petitjean, S. (2015). Implementing semantic frames as typed feature structures with XMG. *Journal of Language Modelling*, 3(1), 185–228.
- Lieber, R. (2004). *Morphology and lexical semantics*. Cambridge: Cambridge University Press.
- Lieber, R. (2016). *English nouns: The ecology of nominalization*. Cambridge: Cambridge University Press.
- Lieber, R., & Andreou, M. (2018). Aspect and modality in the interpretation of deverbal -er nominals in English. *Morphology*, 28(2), 187–217.
- Löbner, S. (2013). *Understanding semantics*. London: Arnold, 2nd, revised edition.
- Löbner, S. (2014). Evidence for frames from human language. In Gamerschlag, T., Gerland, D., Osswald, R. & Petersen, W. (Eds.), *Frames and concept types: Applications in language, cognition, and philosophy* pp. 23–67. Dordrecht: Springer.
- Löbner, S. (2015). Functional concepts and frames. In Gamerschlag, T., Gerland, D., Osswald, R., & Petersen, W., Eds., *Meaning, frames, and conceptual representation*, pp. 35–62. Düsseldorf: dup.
- Perrier, G. (2007). A French interaction grammar. In Angelova, G., Bontcheva, K., Mitkov, R., Nicolov, N. & Simov, K., (Eds.), *International Conference on Recent Advances in Natural Language Processing - RANLP 2007*, pp. 463–467, Borovets, Bulgaria: IPP & BAS & ACL-Bulgaria INCOMA Ltd, Shoumen, Bulgaria.
- Petersen, W. (2007). Representation of concepts as frames. In Skilters, J. (Ed.) *The baltic international yearbook of cognition, logic and communication*, Vol. 2, pp. 151–170.
- Plag, I., Andreou, M., & Kawaletz, L. (2018). A frame-semantic approach to polysemy in affixation. In Bonami, O., Boyè, G., Dal, G., Giraudo, H., & Namer, F. (Eds.), *The lexeme in descriptive and theoretical morphology*. Berlin: Language Science Press.
- Pollard, C., & Sag, I. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press, Stanford : CSLI Publications.
- Rainer, F. (2014). Polysemy in derivation. In Lieber, R. & Štekauer, P. (Eds.), *The oxford handbook of derivational morphology, oxford handbooks in linguistics*, pp. 338–353. Oxford: Oxford University Press.
- Rappaport Hovav M., & Levin, B. (2008). The english dative alternation: The case for verb sensitivity. *Journal of Linguistics*, 44, 129–167.
- Riehemann, S. Z. (1998). Type-based derivational morphology. *Journal of Comparative Germanic Linguistics*, 2(1), 49–77.

- Sag, I. A. (2012). Sign-based construction grammar: An informal synopsis. In Boas, H. C. & Sag, I. A. (Eds.), *Sign-based construction grammar*, pp. 69–202. Stanford, CA: CSLI Publications.
- Schuler, K. K. (2006). *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. PhD thesis, University of Pennsylvania.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Operationalizing the Role of Context in Language Variation: The Role of *Perspective Alignment* in the Spanish Imperfective Domain



Martín Fuchs, María Mercedes Piñango, and Ashwini Deo

**Abstract** We present a cognitively grounded analysis of the pattern of variation that underlies the use of two aspectual markers in Spanish (the Simple-Present marker, *Ana baila* ‘Ana dances’, and the Present-Progressive marker, *Ana está bailando* ‘Ana is dancing’) when they express an *event-in-progress* reading. This analysis is centered around one fundamental communicative goal, which we term *perspective alignment*: the bringing of the hearer’s perspective closer to that of the speaker. *Perspective alignment* optimizes the tension between two nonlinguistic constraints: Theory of Mind, which gives rise to *linguistic expressivity*, and Common Ground, which gives rise to *linguistic economy*. We propose that, linguistically, *perspective alignment* capitalizes on lexicalized meanings, such as the *progressive* meaning, that can bring the hearer to the “here and now”. In Spanish, *progressive* meaning can be conveyed with the Present-Progressive marker regardless of context. By contrast, if the Simple-Present marker is used for that purpose, it must be in a context of shared perceptual access between speaker and hearer; precisely, a condition that establishes *perspective alignment* non-linguistically. Support for this analysis comes from a previously observed yet unexplained pattern of contextually-determined variation for the use of the Simple-Present marker in Iberian and Rioplatense (vs. Mexican) Spanish—in contrast to the preference across all three varieties for the use of the Present-Progressive marker—to express an *event-in-progress* reading.

**Keywords** Imperfective · Progressive · Expressivity · Economy · Common Ground · Theory of mind · Meaning variation and change

---

M. Fuchs · M. M. Piñango

Department of Linguistics, Yale University, New Haven, USA

*Present Address:*

M. Fuchs (✉)

Utrecht Institute of Linguistics - OTS, Utrecht University, Trans 10, Utrecht 3512JK,  
The Netherlands

e-mail: [m.fuchs@uu.nl](mailto:m.fuchs@uu.nl)

A. Deo

Department of Linguistics, The Ohio State University, Columbus, USA

## 1 Introduction

Successful linguistic communication occurs when a speaker utters an expression and a comprehender recognizes the specific meaning that the speaker intended to convey by uttering that expression. If all markers in a linguistic system were in a strict one-to-one correspondence to a meaning, linguistic communication would always be unambiguous. However, that is rarely the case; linguistic markers usually make more than one type of contribution to the composed sentential meaning, leading to different readings of the expressions of which they are part. That is because the markers' associated meanings are encoded in such a way that they demand interaction with a context in order to be properly composed with the other meanings in the expression (e.g., Lewis 1980; Kaplan 1989).

From a communicative perspective, the interaction between linguistic meaning and nonlinguistic context is manifested as a tension between how much meaning is predictably associated with a marker (i.e., lexicalized) and how much meaning must be retrieved from the contextual information in the communicative situation. While the former leads to *expressivity*—the requirement that all intended meaning be linguistically encoded—, the latter leads to *economy*—the possibility that meaning be inferred from the shared history of the interlocutors and the properties of the physical environment where communication takes place at a given time. This tension appears to be rooted in fundamental human cognitive biases: on the one hand, speakers want to be able to convey specific meanings to their hearers; on the other hand, they want to do so by uttering the least amount of linguistic information, relying instead on the contextual properties that constrain the hearer's interpretation. *How are lexical meanings structured such that this tension is resolved, leading to the fast-paced, seemingly transparent, communication process that is typically observed?*

We propose that this question can be addressed by investigating meaning variation; that is, the systematic ways in which a marker shifts its connection to a meaning across members of the same speech community. We hypothesize here that meaning variation for a given marker ultimately results from specific *communicative* and *cognitive* pressures in interaction with the contextual demands of that marker. We focus on *grammatical aspect*, a component of the grammar that is subject to variation and ultimately diachronic change (Dahl 1985; Bybee et al. 1994, *i.a.*); specifically, on the *Imperfective* aspectual domain in Spanish.

The Spanish *Imperfective* aspectual domain is a good test case for analyzing the properties that determine meaning variation given that it is expressed by the Present-Progressive marker and the Simple-Present marker, two markers that convey two readings—the *event-in-progress* and the *habitual*—in a two-by-two system.<sup>1,2</sup>

---

<sup>1</sup>In this paper, we explore the *Imperfective* domain in the Present tense, but we assume that the conclusions that we put forth also hold in a similar way for the *imperfective* and *progressive* meanings in the Past and Future tenses.

<sup>2</sup>These markers are also able to express a *continuous* reading when they are combined with lexically stative predicates, such as in *Ana vive en Bogotá* ('Ana lives in Bogotá') or as in *Ana está viviendo en Bogotá* ('Ana is living in Bogotá'). We leave this reading aside for the purposes of this paper.

The alternations between these two markers also manifest a shared semantic structure between the two meanings that participate in this aspectual domain, in which the *progressive* meaning is a subcase of the more general *imperfective* meaning (Kurylowicz 1964; Comrie 1976; Deo 2009, *i.a.*).

In previous work (Fuchs et al. 2020) we have shown that in Spanish, contrary to traditional assumptions (e.g., Marchand 1955; Bertinetto 2000), these two markers are not in free variation, and that when it comes to the expression of the *event-in-progress* reading, their use appears to be governed by contextual constraints. Here, we present a theoretical model of that variability that is cognitively rooted in the communicative factors involved in those contextual constraints and in the structure of the subsystem(s) to which those communicative factors belong. This model gives rise to an account whereby the recognition of a *progressive* meaning implicates the alignment of the hearer's perspective to that of the speaker. We argue that this alignment can be obtained both by linguistic and by non-linguistic means, and we show that the tension between the use of the Present-Progressive marker and the Simple-Present marker in Spanish to convey an *event-in-progress* reading is a direct result of whether the alignment of the speaker's and the hearer's perspectives was already introduced by non-linguistic means, or whether it needs to be encoded linguistically.

The remainder of this paper is structured as follows. Section 2 describes the distribution of the Present-Progressive marker and the Simple-Present marker in Modern Spanish. Section 3 presents the formal structures we are assuming for the *progressive* and the *imperfective* meanings, together with their communicative implications, and a proposal for a unified meaning structure of these two meanings that allows for the observed systematic variation in their use. Section 4 presents the previously reported data in Fuchs et al. (2020) on the markers' context-modulated behavior in three Spanish varieties for the *event-in-progress* reading. Section 5 presents the analysis based on the data introduced in §4. Section 6 concludes the paper.

## 2 On the Spanish Present-Progressive and Simple-Present Markers

Spanish has two markers that express the *Imperfective* aspectual domain in the Present: the periphrastic Present-Progressive marker in (1a), constituted by the verb *estar* 'to be' plus the gerund *V + -ndo*, and the syncretic Simple-Present marker in (1b) (Yllera 1999; NGRAE 2009, *i.a.*).

- (1) a. Ana est-á fum-ando (ahora).  
Ana be-PRS.3.SG smoke-PROG (now)  
'Ana is smoking now'
- b. Ana fum-a ahora.  
Ana smoke-PRS.3.SG now  
'Ana is smoking now'

In (1) these markers are supporting an *event-in-progress* reading; that is, their contribution to the sentential meaning leads to the interpretation that the event described by the predicate is unfolding at reference time. However, both of these markers can also convey a more general *imperfective* meaning, that, for instance, can give rise to a *habitual* reading; that is, their contribution to the sentential meaning leads to the interpretation that the event described by the predicate has regular instantiations over some interval of time, as in (2).

- (2) a. Ana est-á fum-ando todos los días.  
          Ana be-PRS.3.SG smoke-PROG all the days.  
          ‘Ana is smoking every day’
- b. Ana fum-a todos los días.  
          Ana smoke-PRS.3.SG all the days  
          ‘Ana smokes every day’

The sentences in (1) and (2) show that, given different discourse or situational contexts, both the Present-Progressive marker and the Simple-Present marker can each alternatively convey an *event-in-progress* or a *habitual* reading. This situation raises at least two questions: (1) How are these different readings connected such that this alternation can obtain? (2) If contextual constraints are involved in the observed distribution of the markers, what specific contextual factors are modulating the variation? The answer to these questions is the focus of the next two sections.

### 3 The Meaning of the *Progressive* and the *Imperfective*: A Communicative Perspective

Aspect is said to be the grammatical category that expresses *how* a situation extends over time; from a communicative viewpoint, we can conceive it as a part of the way in which speakers and hearers experience and schematize the world. This experience gets encoded in linguistic devices both lexically and grammatically (e.g., Vendler 1957; Verkuyl 1972; Comrie 1976).

Imperfective aspect denotes a property of a situation whereby the situation is understood as continuing throughout some interval of time. In language-neutral terms, for a sentence to have imperfective aspect, it necessarily and sufficiently needs to present the Subinterval Property; that is, if a predicate *P* is true at some interval *I*, it follows that the predicate *P* is true at all (relevant) subintervals of *I*.

Both the *event-in-progress* and the *habitual* readings of the Spanish *Imperfective* aspectual domain show the Subinterval Property. The sentence radical (smoke(Ana)) in both sentences in (1), repeated here as (3), holds of every relevant subinterval of the reference interval (i.e., *now* in those sentences).<sup>3</sup> In the case of the sentences in

---

<sup>3</sup>We understand *sentence radicals* to be predicates of eventualities with their arguments saturated.

(2), repeated here as (4), the sentence radical in both sentences holds at all relevant regular subintervals of the interval under consideration, which is a superinterval of the reference interval.

- (3) a. Ana est-á fum-ando (ahora).

Ana be-PRES.3.SG smoke-PROG (now)

'Ana is smoking now'

- b. Ana fum-a ahora.

Ana smoke-PRES.3.SG now

'Ana is smoking now'

- (4) a. Ana est-á fum-ando todos los días.

Ana be-PRES.3.SG smoke-PROG all the days.

'Ana is smoking every day'

- b. Ana fum-a todos los días.

Ana smoke-PRES.3.SG all the days

'Ana smokes every day'

Deo (2009, 2015) provides a unified account of the *progressive* and the *imperfective* meanings that allows for the availability of the *event-in-progress* and *habitual* readings. Under this account, the *progressive* and the *imperfective* meanings are encoded as two distinct operators that apply to predicates of eventualities denoted by sentence radicals. This proposal treats the meaning of the *progressive* operator as a subset of the meaning of the *imperfective* operator (see also Kurylowicz 1964; Comrie 1976, *i.a.*). Both operators involve a universal quantifier whose domain of quantification is a *regular partition* of an interval; i.e., a set of collectively exhaustive, non-overlapping, equimeasured subsets of some set, against which the instantiation of a given predicate is evaluated regarding its distribution over time. The notion of instantiation of a predicate over *regular partitions* of an interval captures the intuition of a *regular distribution* over time that obtains with utterances with imperfective aspect. Key to this analysis is that the measure of the regular partition, which determines the value of each cell of the partition, is a free variable with a contextually-determined value. The different readings that each meaning presents are thus the result of different values in different contexts.

The contrast between the two operators emerges from differences in their respective domains of quantification: while in the case of the *progressive* operator, the domain of quantification is a regular partition of the reference interval (that is, the predicate stands in a coincidence relation<sup>4</sup> with regular subintervals of the reference

---

<sup>4</sup>The coincidence relation is defined as follows: "a predicate of events stands in the coincidence relation with an interval  $i$  and a world  $w$  if and only if  $P$  is instantiated in every inertial alternative of  $w$  within  $i$  or at some superinterval of  $i$ " (Deo 2015: 11). Inertia worlds are understood as in Dowty (1977); i.e., as the worlds that continue beyond  $i$  in ways that are compatible with the regular course of events until  $i$ . Inertia worlds thus allow the coincidence relation to avoid the Imperfective

interval), in the case of the *imperfective* operator, the domain of quantification is a regular partition of a *superinterval* of the reference interval (that is, the predicate stands in a coincidence relation with regular subintervals of a superinterval of the reference interval). Thus, the *progressive* meaning behaves as a subset of the *imperfective* meaning: the reference interval is always a subinterval of a superinterval of itself. The formal representations for each of these operators, taken from Deo (2015), are given below:

$$PROG : \lambda P \lambda i \lambda w. \forall j [j \in \mathcal{R}_i^c \rightarrow \text{COIN}(P, j, w)]$$

$$IMPF : \lambda P \lambda i \lambda w. \exists j [i \subseteq_{ini} j \wedge \forall k [k \in \mathcal{R}_j^c \rightarrow \text{COIN}(P, k, w)]]$$

The *progressive* operator combines with a predicate of eventualities  $P$  and an interval  $i$  and returns the proposition that every cell  $j$  of a regular partition of  $i$  coincides with  $P$ . The *imperfective* operator, on the other hand, combines with a predicate of eventualities  $P$  and an interval  $i$ , and returns the proposition that there is some interval  $j$  that continues  $i$  such that every cell  $k$  of a regular partition of  $j$  coincides with  $P$ .

Here we argue that the subset organization dependent on the relation between a reference interval and a superinterval thereof has communicative implications that are observable in specific usage patterns, such as the ones described in §2. Specifically, we propose that the interval structure that underlies both operators constitutes a unified conceptual structure whose variables are the interval under consideration and the measure of the regular partition.<sup>5</sup> The interactions between these two variables give rise to the *event-in-progress* or the *habitual* readings of the different meanings. In what follows, we discuss each meaning and their communicative implementations.<sup>6</sup>

In the case of the *progressive*, the domain of quantification is the reference interval. When the hearer comprehends a *progressive* sentence with an *event-in-progress* reading, such as the sentences in (1), the marker triggers the representation of an interval, the reference interval, as we see in Fig. 1.

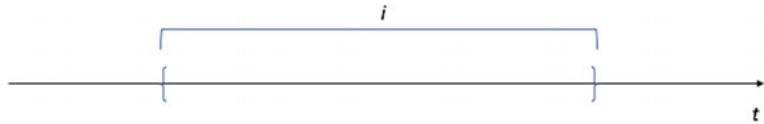
This interval is constituted by regular partitions, as we observe in Fig. 2. What the operator demands is that every cell  $j$  be of a regular partition of  $i$ .

At this point, what is left for the hearer's parser is to map the associated proposition  $P$  to every cell  $j$  of a regular partition of that interval  $i$  in that world of evaluation  $w$ , making it coincide with them, as it can be seen in the visual representation and the formula in Fig. 3.

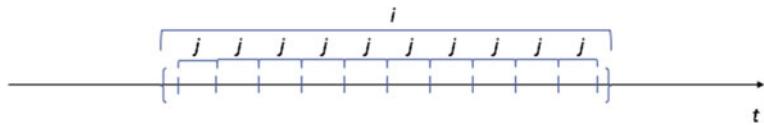
Paradox. Throughout the remainder of the paper, this is the definition of the coincidence relation assumed. We simplify its presentation for reasons of space.

<sup>5</sup>The status of a ‘conceptual structure’ for this meaning structure manifests our deeper claim that this unified meaning is not a linguistic device, but a substructure of a larger nonlinguistic cognitive system to which language has access through imperfective and progressive markers.

<sup>6</sup>The incremental presentations of the communicative implementations of the meanings of the *progressive* and the *imperfective* are not a claim about their processing. They are simply visual devices that illustrate the meaning structure to which the markers have access.

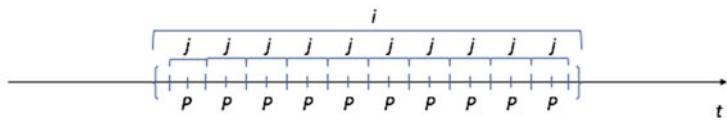


**Fig. 1** The *progressive* meaning from a communicative perspective (1/3)



**Fig. 2** The *progressive* meaning from a communicative perspective (2/3)

$$PROG : \lambda P \lambda i \lambda w. \forall j [j \in \mathcal{R}_i^c \rightarrow \text{COIN}(P, j, w)]$$

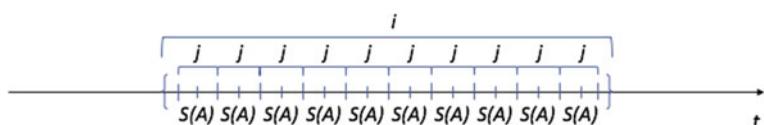


**Fig. 3** The *progressive* meaning from a communicative perspective (3/3)

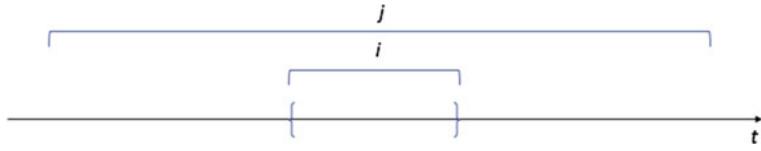
Therefore, a sentence such as (1a), *Ana está fumando ahora*, ‘Ana is smoking now’, would be represented from a communicative perspective as in Fig. 4, where the sentence radical (*Smoke(Ana)*), (*S(A)*), is mapped to every cell of a regular partition of the reference interval.

In the case of the *imperfective*, the domain of the quantifier is a superinterval of the reference interval. This allows for the appearance of the *habitual* reading. From the perspective of communication, when a hearer receives an *imperfective* sentence with a *habitual* reading, it not only triggers the representation of an interval *i*—the reference interval—but also of the associated superinterval *j*, as it can be seen in Fig. 5.

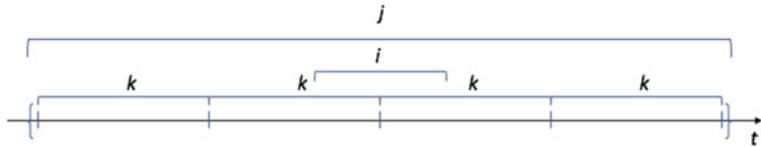
Just like the reference interval, this superinterval is constituted by regular partitions, as we observe in Fig. 6. What the operator demands is that every cell *k* be of a regular partition of *j*.



**Fig. 4** The representation of *Ana está fumando ahora* ‘Ana is smoking now’ from a communicative perspective



**Fig. 5** The *imperfective* meaning from a communicative perspective (1/3)

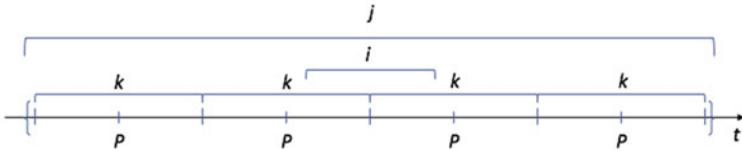


**Fig. 6** The *imperfective* meaning from a communicative perspective (2/3)

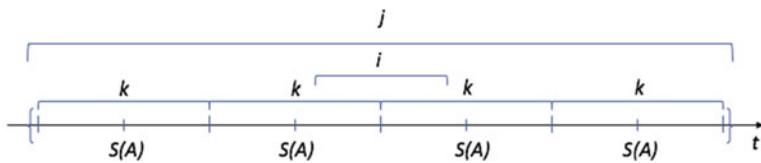
The role of the hearer's parser in this case is to map the proposition  $P$  to every cell  $k$  of a regular partition of that superinterval  $j$  in that world of evaluation  $w$ , making it coincide with them. This is presented in Fig. 7.

Accordingly, from a communicative perspective, a sentence such as (2b), *Ana fuma todos los días*, 'Ana smokes every day', is represented as in Fig. 8. In this case, the sentence radical (*Smoke* (Ana)), ( $S(A)$ ), is mapped to every cell  $k$  of a regular partition of  $j$ .

$$\text{IMPF} : \lambda P \lambda i \lambda w. \exists j [i \subseteq_{\text{ini}} j \wedge \forall k [k \in \mathcal{R}_j^c \rightarrow \text{COIN}(p, k, w)]]$$

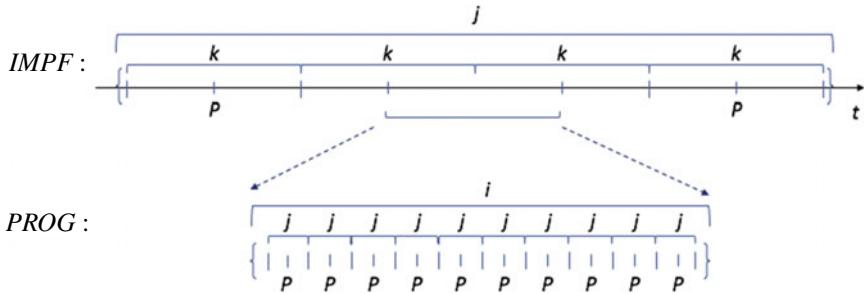


**Fig. 7** The *imperfective* meaning from a communicative perspective (3/3)



**Fig. 8** The representation of *Ana fuma todos los días* 'Ana smokes every day' from a communicative perspective

$$IMPF : \lambda P \lambda i \lambda w. \exists j [i \subseteq_{ini} j \wedge \forall k [k \in \mathcal{R}_i^c \rightarrow \text{COIN}(P, k, w)]]$$

$$PROG : \lambda P \lambda i \lambda w. \forall j [j \in \mathcal{R}_i^c \rightarrow \text{COIN}(P, j, w)]$$


**Fig. 9** The meaning structure of the *imperfective* domain: the *imperfective* (above) and the *progressive* (below)

In Fig. 9 below, these two readings of the *Imperfective* aspectual domain—the *event-in-progress* and the *habitual*—emerge from the same meaning structure: a predicate of events coincides with every cell of a regular partition of an interval. They differ only in the components of the meaning structure that each reading makes salient: while the *habitual* reading makes salient both levels within the structure (the reference interval *and* a superinterval thereof), the *event-in-progress* reading makes salient the reference interval alone.

#### 4 The Markers of the Spanish *Progressive* Are not in Free Variation: Implications

In previous work, we report experimental evidence consistent with the possibility that the Present-Progressive and the Simple-Present markers are not in free variation when conveying an *event-in-progress* reading, and that the choice of marker is in fact contextually determined (Fuchs et al. 2020). In this section, we summarize those results. The data pattern that is presented in that paper serves as a clear test case for our communicative analysis and for testing the implications of a unified conceptual structure for both the *progressive* and the *imperfective* meanings of the *Imperfective* aspectual domain.

Fuchs et al. (2020) reports data from a sentence acceptability judgment task. A total of 114 participants from three different Spanish dialectal varieties rated on a 1-to-5 Likert scale context-sentence pairs that induced an *event-in-progress* reading with either the Present-Progressive marker, the Simple-Present marker, or the Simple-Past marker (used as a baseline condition). Target sentences were preceded either by a context that indicated that speaker and hearer had equivalent perceptual access to the event described by the predicate (Rich Context) or by a context that indicated

that the speaker and the hearer did not share perceptual access to the event (Poor Context). *Shared perceptual access* was operationalized as *visual* perceptual access: both participants in the discourse situation were observing the event that the predicate in the target sentence described. An example of each type of context is presented in (5) and (6) respectively.

### **Rich Context**

- (5) *Ana llega a su casa de trabajar y va a buscar a su hijo a su habitación. Golpea la puerta, la abre, y ve al hijo sentado en el escritorio. Antes de que ella diga nada, el hijo le dice:*

‘Ana comes home from work, and goes to her son’s room to look for him. She knocks on the door, opens it, and sees him sitting at his desk. Before she says anything, her son tells her:’

### **Poor Context**

- (6) *Ana llega a su casa de trabajar y va a buscar a su hijo a su habitación. Golpea la puerta, pero el hijo no contesta. Sin que ella llegue a abrir la puerta, el hijo le dice:*

‘Ana comes home from work, and goes to her son’s room to look for him. She knocks on the door, but her son does not answer. Before she gets to open the door, her son tells her:’

Each of these contexts was then followed by a target sentence that the participant had to rate, which presented either the Present-Progressive marker (7a), the Simple-Present marker (7b), or the Simple-Past marker (7c).

- (7) a. Est-oy                haci-endo                la                tarea.  
           be-PRS.1.SG    do-PROG                the                homework  
           ‘I am doing homework’
- b. Hag-o                    la                tarea.  
           do-PRS.1.SG    the                homework  
           ‘I am doing homework’
- c. Hi-ce                    la                tarea.  
           do-PST.1.SG    the                homework  
           ‘I did homework’

The study was originally designed to test two competing hypotheses regarding the variation between these markers to express an *event-in-progress* reading: a *free alternation hypothesis*, which argued that the markers could be used interchangeably regardless of the type of context, and a *context dependent hypothesis*, which stated

**Table 1** Participants' ratings means and standard errors by condition (dialect \* aspectual marker \* context)

	Iberian Spanish		Rioplatense Spanish		Mexican Alt. Spanish	
	Rich	Poor	Rich	Poor	Rich	Poor
P. PROG	4.78 (0.03)	4.74 (0.03)	4.68 (0.05)	4.66 (0.05)	4.51 (0.06)	4.46 (0.06)
S. PRESENT	<b>4.18 (0.11)</b>	<b>3.70 (0.09)</b>	<b>3.90 (0.11)</b>	<b>3.43 (0.08)</b>	3.57 (0.12)	3.51 (0.12)
S. PAST	2.16 (0.08)	2.15 (0.09)	2.67 (0.08)	2.57 (0.08)	2.67 (0.09)	2.63 (0.08)

that the choice of marker was conditioned by properties of the contextual information. We proposed that marker use was context-dependent, and that its locus of variation was *shared perceptual access* to the event between the speaker and the hearer.<sup>7</sup> The three varieties of Spanish probed were Mexican Altiplano Spanish (Mexico City), Iberian Spanish (Madrid), and Rioplatense Spanish (Buenos Aires) with similar participant distributions: 39 (20 female) Iberian Spanish speakers; 38 (21 female) Rioplatense Spanish speakers, and 37 (21 female) Mexican Altiplano Spanish speakers.<sup>8</sup> The rationale for testing different varieties was that the *Imperfective* aspectual domain could be partitioned by these markers in different yet predictable ways in each of the dialects.

A summary of the results in terms of the participants' ratings means by context, aspectual marker and dialect is given in Table 1. Standard errors are indicated in parentheses. Conditions where there are significant differences are bolded.

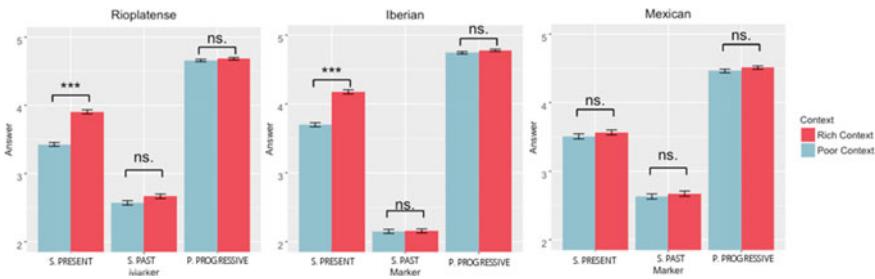
In all three Spanish varieties, the Present-Progressive marker is the preferred form to express an *event-in-progress* reading regardless of contextual information, while the Simple-Past form is disallowed from expressing an *event-in-progress* reading

<sup>7</sup>With respect to the Simple-Present marker, we tested the prediction associated with the *context dependent hypothesis*. According to this hypothesis, when the situational context presents information that shows that speaker and hearer share perceptual access to the event described by the predicate, the Simple-Present marker should get significantly higher ratings than when the information in the situational context does not indicate that speaker and hearer share perceptual access to the situation at issue.

Regardless of the issue of context-dependence, we expected the Present-Progressive marker in every dialect to obtain ceiling ratings, as the Present-Progressive marker exhibits the *event-in-progress* as its most salient reading. Our analysis argued that this occurred because the Present-Progressive marker was unambiguous in conveying an *event-in-progress* reading. That analysis, however, was incomplete in that it did not take into account the *habitual* reading of the Present-Progressive marker, such as the one in (2a), *Ana está fumando todos los días* ‘Ana is smoking every day’, whose existence evidences that the locus of the variation is not necessarily presence/absence of ambiguity in marker-meaning correspondence, but something else that relates the structure of the meaning itself (i.e., the *progressive*) to its communicative implications.

The model we present here accounts for the presence of ambiguity by arguing that while the Present-Progressive marker may be preferentially lexically associated with the *progressive* meaning, given the shared conceptual structure described in §3, it also has the potential to access the other readings. It does so by allowing modification of the measure of the regular partition—and, in this way, referring to a superinterval of the reference interval—, thus achieving a *habitual* interpretation. Unfortunately, more extensive discussion of these cases is beyond the scope of this paper.

<sup>8</sup>For details on the procedure, see Fuchs et al. (2020), §4.2.



**Fig. 10** Participants' means by context condition, aspectual marker and dialect

across the board. With respect to the Simple-Present marker, in at least Rioplatense and Iberian Spanish, the acceptability of the marker appears to be modulated by contextual information. When the speaker and the hearer share perceptual access to the event described by the predicate, participants judge the use of the Simple-Present marker as significantly more acceptable than when the speaker and the hearer do not share perceptual access to the event. In the case of Mexican Altiplano Spanish, the Simple-Present marker is dispreferred with respect to the Present-Progressive marker regardless of contextual information.<sup>9</sup> A graph of the participants' ratings by contextual information, marker and dialect is presented in Fig. 10.

These results show that the use of the Simple-Present marker to convey an *event-in-progress* reading is restricted by context in at least two dialects of Spanish—Rioplatense and Iberian Spanish. Therefore, the data show that the markers do not alternate freely, and provide support to the *context dependent hypothesis*. While the Present-Progressive marker is the preferred form to convey an *event-in-progress* reading across the three dialectal varieties and regardless of contextual information, the Simple-Present marker is context-dependent and its acceptability is modulated by the assessment that participants make of the *shared perceptual access* between speaker and hearer conveyed in the preceding context. We also observe that this context-dependence is subject to dialectal variation: while Rioplatense and Iberian Spanish show context-dependence in their use of the Simple-Present marker, Mexican Altiplano Spanish presents a distribution in which this contextual distinction becomes irrelevant, and the only mean to achieve the *event-in-progress* reading is linguistic; that is, the use of the Present-Progressive marker.

<sup>9</sup>For a detailed explanation of why the dialects differ, and how this variation is constrained by a unidirectional diachronic grammaticalization path from Progressive to Imperfective, see Fuchs et al. (2020), §2.2., and §6.

## 5 Analysis: The Psychological Roots of *Shared Perceptual Access*

The pattern described in §4 shows that the distribution between the Present-Progressive marker and the Simple-Present marker in the expression of the *event-in-progress* reading is not haphazard, but governed by contextual constraints; namely, by whether the speaker and the hearer *share perceptual access* to the event described by the predicate.

In this section, we present an analysis of this contextual factor that is couched in terms of general communicative and cognitive constraints. Our proposal is based on the notion of *perspective*, understood as the information that is perceptually available for a given individual from a particular point of view in space (Roberts 2015: 3). This perspective, moreover, is *doxastic* in that it is understood to be the set of worlds compatible with an individual's beliefs at that time in that world. From a communicative perspective, we consider that grammatical aspect not only reflects the point of view of the speaker, but it is also able to manipulate it, in a process that we call *perspective alignment*. In this process, which we consider to be one of the general goals of communication, the speaker intends to align the hearer's (doxastic) perspective to her own; that is, she intends to make the worlds compatible with the hearer's beliefs more like the worlds compatible with her own beliefs.

We propose *perspective alignment* as the resolution of the well-known tension between linguistic *economy* and linguistic *expressivity* during communication (Zipf 1949). We take these two factors to be epiphenomenal: manifestations of different kinds of knowledge. On the one hand, *linguistic economy* reflects a speaker's expectation about the hearer that, given their shared history, their minds' perception and schematization of the world are the same. This expectation allows the speaker to make her utterances shorter, containing more lexical items with underspecified meanings. Linguistic *economy* is thus a manifestation of the Common Ground, the *shared context* between interlocutors during a given linguistic communicative act (Stalnaker 1978, 2002; Roberts 1996/2012 *i.a.*). It is the speaker's expected common ground with the hearer that allows for linguistic economy.

Linguistic *expressivity*, on the other hand, reflects the speaker's knowledge that the hearer is a separate individual and that consequently their minds may overlap but are not identical and are not necessarily experiencing and schematizing the context at issue in the same manner. From a linguistic communicative perspective, this knowledge amounts to Theory of Mind (Wellman 1990; Gopnik 1993; de Villiers 2007, *i.a.*). This understanding compels the speaker to encode linguistically all of her intended meaning, leading to linguistic *expressivity*.

Under these two notions, linguistic *economy* appears as speaker-oriented, while linguistic *expressivity* appears as hearer-oriented. Thereby lies the communicative tension that clarifies the objective of linguistic communication: the bringing of the hearer to the point of view or perspective of the speaker. And this, in a nutshell, is what *perspective alignment* seeks: the optimization of Common Ground and Theory of Mind constraints between speaker and hearer during the communicative act.

We argue that, linguistically, *perspective alignment* can be achieved by lexicalized meanings, such as the *progressive* meaning, that bring the hearer to the “here and now”. The *progressive* meaning makes salient the reference interval in the shared meaning structure (described in §3)—thus conveying information about the “here and now”—, and in doing so, it brings the perspective of the hearer closer to that of the speaker.

Under this analysis, when intending to convey a *progressive* meaning in a language with two distinct markers whose alternation is contextually determined—such as present-day Spanish—, the speaker has either the choice of relying on non-linguistic contextual information and use the Simple-Present marker or the choice of using the Present-Progressive marker. In order to felicitously utter a sentence with a Simple-Present marker that conveys a *progressive* meaning, the speaker needs to know that the hearer has perceptual access to the situation described by the embedded proposition. This condition—shared perceptual access—constraints the interpretation to the reference interval, satisfying the requirements of the *progressive* meaning, and brings about *perspective alignment* by non-linguistic means. If the speaker cannot know whether the hearer has perceptual access to the situation described by the embedded proposition, *perspective alignment* is not met non-linguistically, and the Present-Progressive marker must be used instead. In this way, *perspective alignment* can be provided both non-linguistically (by contextual information) or linguistically (by the use of the Present-Progressive marker).

This is what the pattern uncovered in Fuchs et al. (2020) ultimately shows: that the acceptability of the Simple-Present marker to convey a *progressive* meaning increases in Rioplatense and Iberian Spanish, but only when the situational context expresses that there is shared perceptual access to the event between speaker and hearer, guaranteeing non-linguistically speaker-hearer *perspective alignment*. Conversely, in cases in which the information given in the situational context does not indicate that there is shared perceptual access to the event between speaker and hearer, and *perspective alignment* is not provided non-linguistically, the acceptability of the Simple-Present marker decreases significantly. In these cases, the speaker needs to assume that the hearer can only rely on linguistic information to comprehend the intended meaning that she wants to convey, and resort to the Present-Progressive marker. In sum, the Simple-Present marker can be used to convey a *progressive* meaning only when the communicative goal of *perspective alignment* is achieved independently.

Finally, even in rich contexts, where *perspective alignment* is non-linguistically guaranteed, we observe that the Present-Progressive marker gets higher ratings than the Simple-Present marker. We account for this pattern by invoking a key property of language: lexicalization as a means to faster processing. The Present-Progressive marker, by its preferred reference interval interpretation (*progressive*), has in a way lexicalized *perspective alignment*.<sup>10</sup> By contrast, the use of the Simple-Present marker to reach *perspective alignment* demands the incorporation of non-linguistic

---

<sup>10</sup>We claim that this is true not only for the sentences in which the Present-Progressive marker conveys a *progressive* meaning, but also for sentences such as (2a), *Ana está fumando todos los días* ‘Ana is smoking every day’, where the Present-Progressive marker does not express an *event-in-progress* reading, but a *habitual* one with a temporal contingency. In these cases, *perspective*

information, which ultimately needs to be integrated into a unified meaning structure. As comprehension progresses, such real-time integration of linguistic and contextual information is arguably computationally costlier. And it is the avoidance of this cost what finally leads speakers to systematically prefer Present-Progressive-marked utterances. An extreme version of this situation is shown by the Mexican Altiplano Spanish variety, in which the Simple-Present marker is dispreferred to convey a *progressive* meaning even when the context provides *perspective alignment* by non-linguistic means.

## 6 Summary and Conclusions

Here we have provided a cognitively grounded approach to non-linguistic context modeling, and an account of how contextual factors interact with linguistic information in the process of sentence meaning comprehension. We have capitalized on a pattern previously reported (Fuchs et al. 2020), which shows that across two varieties of Spanish the acceptability of the Simple-Present marker to convey a *progressive* meaning is modulated by whether or not the speaker and the hearer share perceptual access to the situation described by the proposition at issue.

We have shown that this contextual factor can be captured by appealing to a core communicative goal: *perspective alignment*. This communicative goal is taken to be the optimization of the tension between linguistic *economy*—rooted in Common Ground—and linguistic *expressivity*—rooted in Theory of Mind. The connections with deeper cognitive capacities render *shared perceptual access* not a primitive, but the non-linguistic operationalization of this generalized communicative objective, *perspective alignment*. As the data show, *shared perceptual access* is necessary whenever the linguistic marker cannot bring about *perspective alignment* on its own. Such is the case of the Spanish Simple-Present marker when it is conveying *progressive* meaning. By contrast, when the linguistic marker is the Spanish Present-Progressive marker, it can signal *perspective alignment* on its own. In doing so it presents two communicative advantages: (1) it makes communicative success more predictable, and therefore efficient, since its use is now less context-dependent, and (2) it demands less computational resources: it saves the processor the cost of integrating the linguistic content and the non-linguistic contextual information that it would otherwise need to achieve a felicitous interpretation. These communicative advantages predict in turn an asymmetry in preference between the Simple-Present and the Present-Progressive markers in favor of the latter. This prediction is borne out by the variation pattern: across three Spanish dialectal varieties, the Present-Progressive marker is preferred over the Simple-Present marker to convey the *progressive* meaning regardless of context. This preference is particularly telling

---

*alignment* also obtains even though the *ongoingness* of the event is not at issue; that is, the perspective of the hearer is also brought closer to that of the speaker even if the event is not unfolding at reference time. We leave the analysis of these cases for further research.

in the case of the Mexican Altiplano variety. In this variety, the Simple-Present marker no longer shows context sensitivity effects, suggesting that the Simple-Present marker is no longer able to participate in the achievement of *perspective alignment* even when the main components of this communicative goal are independently (non-linguistically) provided by the shared perceptual access to the event between speaker and hearer. On the assumption that the Mexican Altiplano variety, like the other two varieties, showed these context effects at some previous point in its diachrony, the absence of context effects in the variety's modern instantiation suggests the resolution of a competition for the signaling of *perspective alignment* between the two markers; a competition that the Present-Progressive marker won. As it turns out, such a pattern is not idiosyncratic to Spanish. It is instead consistent with the well-attested cross-linguistic diachronic pattern of encroachment of Present-Progressive markers over the aspectual domain originally covered by Simple-Present markers (e.g., Bybee et al. 1994; Deo 2015).

Altogether, the approach to context structure presented here is consistent with a view of a relation between grammar and meaning that is mediated by generalized nonlinguistic communicative goals, such as *perspective alignment*, that can be lexically harnessed, that are at play during real-time language comprehension, and that link individualized usage patterns with the behavior of dialectal varieties and with generalized cross-linguistic patterns of change.

**Acknowledgements** We would like to thank both anonymous reviewers, and the audiences at HLS 2016, SNEWS 2016, Cognitive Structures 2016, CUNY 2017 and ICHL 2017 for comments and discussion about this work. All errors and omissions are our own. This research has been funded by NSF-INSPIRE Grant CCF-1248100, “The Underpinnings of Semantic Change: A Linguistic, Cognitive, and Information-Theoretic Investigation” to María Mercedes Piñango, Todd Constable, Ashwini Deo, and Mokshay Madiman.

## References

- Bertinetto, Pier Marco. (2000). The progressive in Romance, as compared with English. In Ö. Dahl (Ed.), *Tense and aspect in the languages of Europe*. (pp. 559–604). Berlin: de Gruyter.
- Bybee, Joan, Perkins, Revere, & Pagliuca, William. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: UCP.
- Comrie, Bernard. (1976). *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.
- Dahl, Östen. (1985). *Tense and Aspect Systems*. Oxford: Basil Blackwell.
- de Villiers, Jill. (2007). The interface of language and theory of mind. *Lingua*, 117, 1858–1878.
- Deo, Ashwini. (2009). Unifying the imperfective and the progressive: Partitions as quantificational domains. *Linguistics and Philosophy*, 32, 475–521.
- Deo, Ashwini. (2015). The semantic and pragmatic underpinnings of grammaticalization paths: The progressive to imperfective shift. *Semantics and Pragmatics*, 8, 1–52.
- Dowty, David R. (1977). Towards a semantic analysis of verb aspect and the english ‘imperfective’ progressive. *Linguistics and Philosophy*, 1, 45–77.
- Fuchs, Martín, Deo, Ashwini & Piñango, María Mercedes. (2020). The progressive-to-imperfective shift: contextually determined variation in Rioplatense, Iberian, and Mexican Altiplano Spanish.

- In A. Morales-Front, M. J. Ferreira, R. P. Leow, & C. Sanz (Eds.), *Hispanic Linguistics. Current trends and new directions* (pp. 119–136). Amsterdam: John Benjamins.
- Gopnik, Alison. (1993). How we know our minds: the illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, 16, 1–15.
- Kaplan, David. (1989). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). New York: Oxford University Press.
- Kurylowicz, Jerzy. (1964). *The inflectional categories of Indo-European*. Heidelberg: Carl Winter.
- Lewis, David. (1980). Index, context, and content. In S. Kanger, & S. Öhman (Eds.), *Philosophy and grammar* (pp. 79–100). Dordrecht: D. Reidel Publishing Company.
- Marchand, Hans. (1955). On a question of aspect: A comparison between the progressive form in english and that in Italian and Spanish. *Studia Lingüística*, 9, 45–52.
- NGRAE. Real Academia Española y Asociación de Academias de la Lengua Española. (2009). Las perifrasis verbales. *Nueva Gramática de la Lengua Española*. (pp. 2105–2222). Madrid: Espasa.
- Roberts, Craige. (1996/2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. In J. H. Yoon, & A. Kathol (Eds.), *Papers in Semantics*, 49. Columbus, OH: Ohio State University Department of Linguistics. *Semantics and Pragmatics*, 5(6), 1–69.
- Roberts, Craige. (2015). Indexicality: de se semantics and pragmatics. *Ms. Ohio State University*.
- Stalnaker, Robert. (1978). Assertion. In P. Cole (Ed.), *Pragmatics* (pp. 315–332). New York: New York Academic Press.
- Stalnaker, Robert. (2002). Common ground. *Linguistics and Philosophy*, 25, 701–721.
- Vendler, Zeno. (1957). Verbs and times. *Philosophical Review*, 66(2), 143–160.
- Verkuyl, Henk. (1972). *On the compositional nature of the aspects*. Dordrecht, Netherlands: Reidel.
- Wellman, Henry. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.
- Yllera, Alicia. (1999). Las perifrasis verbales de gerundio y participio. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (pp. 3390–3441). Madrid: Espasa.
- Zipf, George K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley Press.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# A Frame-Based Analysis of Verbal Particles in Hungarian



Kata Balogh and Rainer Osswald

**Abstract** The verbal particle in Hungarian raises a number of intriguing issues for any theory of the syntax-semantics interface. In this article, we aim at a formal account of the semantic contribution of various verbal particles in Hungarian and we show how the semantic representation of the clause can be compositionally derived. We will concentrate on the four frequent particles *meg-*, *le-*, *el-* and *fel-*. Our approach makes use of a formalized version of Role and Reference Grammar and the framework of decompositional frame semantics. In particular, we give a formal representation of the boundary-setting function of the verbal particle in terms of decompositional frames which builds on a scalar change analysis. We furthermore analyze the interaction of the particle with resultative adjectives and provide a formal model of how their syntactic representations drive their frame-semantic composition.

**Keywords** Verbal particles · Hungarian · Scalar change · Decompositional frame semantics · Role and Reference Grammar.

## 1 The Verbal Particle in Hungarian

The verbal particle in Hungarian raises a number of intriguing issues for any theory of the syntax-semantics interface. In its default position immediately preceding the verb (1a), the verbal particle stands in complementary distribution with other verbal modifiers such as resultative predicates (1b), bare nouns and infinitival com-

---

<sup>1</sup> Abbreviations: ACC ‘accusative’, ILL ‘illative’, INESS ‘inessive’, PAST ‘past tense’, PL ‘plural’, POSS ‘possessive’, SUPESS ‘superessive’, SUBL ‘sublative’, VPTCL ‘verbal particle’.

K. Balogh (✉) · R. Osswald  
Institute of Linguistics and Information Science, Heinrich-Heine-Universität, 40204  
Düsseldorf, Germany  
e-mail: [Katalin.Balogh@hhu.de](mailto:Katalin.Balogh@hhu.de)

plements.<sup>1</sup> (Moreover, the immediate preverbal position can host the narrow focus constituent and sentential negation.)

- (1) a. Anna le-festette a kerítés-t.  
Anna VPTCL-paint.PAST the fence-ACC  
'Anna painted the fence.'
- b. Anna zöld-re festette a kerítés-t.  
Anna green-SUBL paint.PAST the fence-ACC  
'Anna painted the fence green.'

Hungarian verbal particles vary considerably with respect to their origin (e.g. Forgács 2004) and their semantic contribution (e.g. Kiefer and Ladányi 2000).<sup>2</sup> Several particles express directionality (e.g. *le-* 'down', *ki-* 'out') while others, including the frequent particle *meg-*, are more difficult to classify on the basis of their lexical meaning. In the following, we will focus on interpretational aspects of the four verbal particles *meg-*, *le-* ('down, off'), *el-* ('away') and *fel-* ('up'), which, together with *ki-* ('out') and *be-* ('in'), constitute the six oldest verbal particles in Hungarian (cf. Szoltész 1959). The overall goal of this article is to give a formal account of the semantic contribution of these particles, and to show how the semantic representation of the clause can be compositionally derived.

In a particle-verb combination, the verbal particle may contribute its original lexical meaning, as, for instance, directionality in the examples in (2), or the particle may have a more abstract semantic effect on the meaning of the verb as in (1a) above.

- (2) a. Anna le-szaladt a pincé-be.  
Anna VPTCL-run.PAST the basement-ILL  
'Anna ran down to the basement.'
- b. Anna hirtelen el-szaladt.  
Anna suddenly VPTCL-run.PAST  
'Anna suddenly ran away.'

The directional meaning is mostly present in combination with verbs of motion as shown in (2). In this case, the verbal particle is often characterized as terminative (Kiefer and Ladányi 2000, pp. 25f; É. Kiss 2008). The example in (1a) illustrates the non-directional meaning contribution of the particle when combined with a non-motion verb such as *fest* ('paint'). In such cases, Kiefer and Ladányi (2000) analyze the verbal particle as a "functor" that changes the Aktionsart of the predicate, e.g., by expressing a boundary condition. The introduction of an end or result condition is a frequent example of Aktionsart formation.

Traditionally, *meg-* has mostly been regarded as a pure aspectualizer or perfectivizer signaling perfective aspect and, thereby, determining the viewpoint aspect, as illustrated by the contrast between (3a) and (4a). In more recent studies, *meg-* is often taken as a delimiter (e.g., Bene 2009), signaling telicity (e.g., Kardos 2016)

---

<sup>2</sup>For more information on the historical development of the verbal particles in Hungarian see e.g. Szoltész (1959) and Pátronics (2002).

and, thus, relating to the Aktionsart (lexical aspect) of the predicate (e.g., Kiefer 2009; Kiefer and Németh 2012). The particle *meg-* is exclusively used in this way (4a), and the other particles discussed in this article can all be used in this way as shown, for example, in (4b) and (4c).

- (3) a. Száradt a törölköző.  
dry.PAST the towel  
'The towel was drying.'
  - b. Anna festette a kerítés-t.  
Anna paint.PAST the fence-ACC  
'Anna was painting the fence.'
  - c. Péter mosta a padló-t.  
Peter wash.PAST the floor-ACC  
'Peter was washing/mopping the floor.'
- (4) a. Meg-száradt a törölköző.  
VPTCL-dry.PAST the towel  
'The towel (has) dried.'
  - b. Anna le-festette a kerítés-t.  
Anna VPTCL-paint.PAST the fence-ACC  
'Anna (has) painted the fence.'
  - c. Péter fel-mosta a padló-t.  
Peter VPTCL-wash.PAST the floor-ACC  
'Peter (has) washed/mopped the floor.'

The choice of the particle seems to be sensitive to the fine-grained semantic class of the base verb, at least to a certain extent. For instance, similar to the case of *le-fest* ('paint sth'), the particle *le-* combines with a number of other verbs which express a surface oriented incremental change such as *le-töröl* ('wipe down'), *le-söpör* ('sweep') and *le-arat* ('harvest'). Moreover, particle verbs of this group can co-occur with a resultative phrase, in which case the verbal particle occupies the preverbal position and the resultative phrase appears postverbally (5).

- (5) Anna le-festette zöld-re a kerítés-t.  
Anna VPTCL-painted green-SUBL the fence-ACC  
'Anna painted the fence green.'

Other classes of verbs, including verbs of creation (e.g. *meg-ír* 'write', *fel-épít* 'build up'), allow for either a particle or a resultative phrase in the preverbal position, but reject the co-occurrence of the two. Yet others, including verbs of performance and perception of performances (e.g. *el-énekel* 'sing', *meg-hallgat* 'listen to'), seem not to allow for a resultative phrase at all.

Irrespective of the fact that the verbal particle affects the Aktionsart (lexical aspect) of the predicate, the syntactic position of the particle can have an influence on the aspectual interpretation (viewpoint aspect) of the utterance. The immediate preverbal

position of the particle is associated with a perfective interpretation. The inverse order, by contrast, gives rise to a progressive interpretation (6).<sup>3</sup>

- (6) Anna (éppen) festette le a kerítést, amikor meg-érkezett Péter.  
 Anna (just) painted VPTCL the fence-ACC when VPTCL-arrived Peter  
 'Anna was painting the fence, when Peter arrived.'

In (6), the presence of the particle still indicates an intended result state while the postverbal position of the particle signals that the viewpoint aspect is progressive.

As mentioned above, the locative or directional meaning component of the particle, if available, is largely restricted to base verbs that denote movements or spatial positions. In these cases, Kiefer and Ladányi (2000) analyze the verbal particle as a predicate of location or direction and É. Kiss (2008) argues that the verbal particle has a terminative role and signals the end position of the moving theme as in (2a). The latter analysis seems problematic in view of examples like (2b), where the particle does not signal a final location or terminativity but the (deictic) direction of the movement.

Another possible function of verbal particles is to signal the inception or inchoation, i.e. the beginning of an event (or state). The particles *meg-*, *el-* ('away') and *fel-* ('up') can contribute this meaning component. Examples are *el-alszik* ('fall asleep'), *meg-szeret* ('get to love') and *fel-zúg* ('begin to buzz'). In (7), the base verb *zúg* ('buzz') denotes the production of a humming sound. The verbal particle *fel-* in (7b) signals the beginning of this activity or process.

- (7) a. Zúg a motor.  
 buzz the engine  
 'The engine is buzzing.'  
 b. Fel-zúg a motor.  
 VPTCL-buzz the engine  
 'The engine starts to buzz.'

Similarly, the particle verbs *el-alszik* ('fall asleep') and *meg-szeret* ('get to love') refer to the inchoation of an activity/state of sleeping and a state of loving. However, these predicates slightly differ from the one in (7b). As Kiefer and Ladányi (2000) point out, both *el-alszik* and *meg-szeret* can be modified by the adverbial *lassan* ('slowly'), cf. (8a) as opposed to *fel-zúg* (8b).

- (8) a. Anna lassan el-aladt.  
 Anna slowly VPTCL-slept  
 'Anna slowly fell a sleep.'  
 b. #Lassan fel-zúg a motor.  
 slowly VPTCL-hum the engine

---

<sup>3</sup>The inverse order can also be triggered by other means: Narrow focus and negation are required to appear in the immediate preverbal position, causing the verbal particle to appear postverbally. In these cases, the viewpoint aspect of the clause remains neutralized or ambiguous.

This suggests that some preparatory phase is present in the case of *el-alszik* and *meg-szeret*. We propose an analysis for both (7b) and (8a) representing inchoation as referring to the initial part of the activity/state contributed by the base verb. The difference in the possibility of adverbial modification of *el-alszik* and *fel-zúg* can be explained by differences in the temporal extension of this initial part.

## 2 Scalar Analysis and Frame-Semantic Representation

In Sect. 3 below, we propose a formal semantic analysis of the data discussed so far that combines a *scalar* analysis of the verbal particle with *frame-based* semantic representations of the lexical items involved. The purpose of the present section is twofold: First, we briefly review the scalar approach, which has been put forward as a general framework for the analysis of aspectual properties in the verbal domain by Filip (2008), Rappaport Hovav (2008), Kennedy and Levin (2008), and Beavers (2008), among others, based on the work of Krifka (1998) and Hay et al. (1999). Second, we introduce decompositional frame semantics as a representational means that integrates frame semantics with lexical decomposition and formal semantics. In particular, we will show how changes along a scale can be represented in frames.

The basic idea of the scalar approach is that gradual changes expressed by verbs or verbal constructions can be uniformly characterized as monotonic changes along an ordered set of degrees with respect to a certain dimension of measurement. Under this analysis, telicity comes about by boundaries on the scale, which can be inherent to the scale or imposed on it by the context. An early focus of the scalar approach was the analysis of deadjectival degree achievements such as *widen* and *dry*. The two verbs differ in that the scale associated with *widen* is open while the one associated with *dry* is closed, which has consequences for their default aspectual interpretation (Kearns 2007).

The scalar viewpoint has been fruitfully applied to the analysis (Kagan 2013, 2016; Zinova 2017). A common assumption of these approaches is that the prefixes determine a dimension of measurement on the basis of a scalar structure given by the base verb and, possibly, its direct, oblique, or prepositional object.

First applications of the scalar approach to the analysis of verbal particles and telicity in Hungarian are given in Kardos (2012, 2016) and Csirmaz (2012). As indicated in the previous section, the distinction between atelic and telic uses of deadjectival degree achievement verbs in Hungarian is marked by the presence of a verbal particle or another boundary-setting element in preverbal position. The contrast between (3a) and (4a) illustrates this for the intransitive verb *szárad* ('dry'), which is related to the adjective *száraz* ('dry'). The simple past tense use without a verbal particle shown in (3a) describes the process of drying, i.e., of getting drier. If the particle *meg-* is added, the resulting verb is telic and describes the accomplishment of getting dry; cf. (4a). This pattern carries over to transitive verbs such as *fest* ('paint') and *mos* ('wash'), which can be used to denote activities as well as accomplishments. When combined with a direct object that encodes a quantized predicate (cf. Krifka

1998), the presence or absence of a verbal particle (or a resultative expression) determines the interpretation as an activity (atelic) or an accomplishment (telic), respectively. This contrast is illustrated in (3b) versus (4b) and in (3c) versus (4c).<sup>4</sup> According to the descriptive analysis of Kardos (2016), the verbal particle (or a resultative expression) encodes an event-maximalization operator in the sense of Filip (2008) which goes along with the presence of a closed scale. That is, the particle in preverbal position imposes a bound on the event denoted by the verb. In the formal representations presented below, this corresponds to the existence of a *final event stage* in which the maximal value of the associated scalar attribute holds at the relevant event participant. For instance, in the final stage of a drying event described as telic, the affected object is characterized as having maximal dryness (or zero moisture).

The formal semantic framework employed in the following makes use of *decompositional frames* (Kallmeyer and Osswald 2013; Osswald and Van Valin 2014). A crucial assumption of frame semantics is that attributes (features, functional relations) play a central role in the organization of semantic and conceptual knowledge and semantic representation (Barsalou 1992; Löbner 2014). Frames are thus inherently structured representations whose semantic components (participants, subevents etc.) can be recursively accessed via attributes. Another aspect of the presented approach is that semantic computation can be understood as the incremental construction of (minimal) frame models based on the input, the context, the lexicon, and background knowledge, while composition is basically realized by frame unification under constraints.

A standard decomposition structure like the one shown in (9) for transitive *break* (cf., e.g., Levin and Rappaport Hovav 2011) can be represented as an event frame of type *causation* which has a CAUSE component of type *activity* and an EFFECT component of type *change-of-state*, which in turn has a RESULT component of type *broken*.

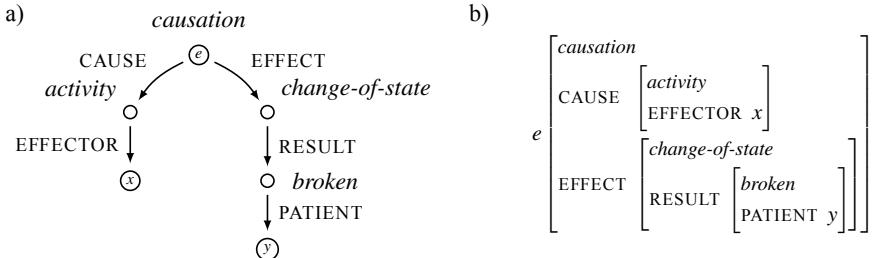
- (9) [[x ACT] CAUSE [BECOME [y BROKEN]]]

Moreover, the participants *x* and *y* are represented as the EFFECTOR of the activity and the PATIENT of the result component, respectively. The overall frame structure is graphically depicted in Fig. 1a.

Formally, we define frame structures as base-labeled feature structures with types and relations as introduced in Kallmeyer and Osswald (2013). Structures of this type arise as canonical models of certain attribute-value descriptions. For example, the frame structure in Fig. 1a is the canonical model of the (closed) attribute-value description in (10).<sup>5</sup> The attribute-value matrix shown in Fig. 1b can be seen as a notational variant of this description.

<sup>4</sup> As noted by Kardos (2016, pp. 4ff, 28ff), verbs of consumption and creation behave somewhat differently in that they may receive a telic interpretation even without a verbal particle if the direct object has quantized reference.

<sup>5</sup> The corresponding *open* (or *unlabelled*) description, which lacks the leading label *e*, can be seen as a one-place predicate that is either true or false at the nodes of a frame structure.



**Fig. 1** Frame representation and attribute-value matrix

$$(10) \quad e \cdot (\text{causation} \wedge \text{CAUSE } \text{activity} \wedge \text{CAUSE } \text{EFFECTOR} \triangleq x \wedge \\ \text{EFFECT } \text{change-of-state} \wedge \text{EFFECT } \text{RESULT } \text{broken} \wedge \\ \text{EFFECT } \text{RESULT } \text{PATIENT} \triangleq y)$$

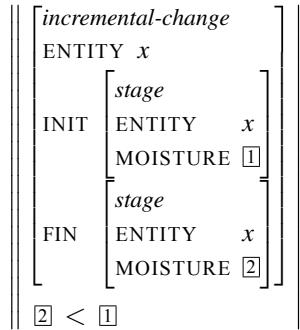
Attribute-value descriptions have a straightforward translation into expressions of first-order predicate logic. The respective translation of (10) is given in (11), with  $e$ ,  $x$  and  $y$  used as free variables (or constants) and with the additional requirement that all attribute relations (written in small caps) are functional.

$$(11) \quad \exists e' \exists e'' \exists s (\text{causation}(e) \wedge \text{CAUSE}(e, e') \wedge \text{EFFECT}(e, e'') \wedge \text{activity}(e') \wedge \\ \text{EFFECTOR}(e', x) \wedge \text{change-of-state}(e'') \wedge \text{RESULT}(e'', s) \wedge \\ \text{broken}(s) \wedge \text{PATIENT}(s, y))$$

The structure in Fig. 1a can then be characterized as the *minimal model* of (11) in the usual sense of first-order predicate logic, under the assumption that attribute relations are functional.

The formal framework just sketched has no direct means to encode universal quantification. In order to be able to represent the implicit quantification over subevents (or subintervals) involved in the characterization of a scalar change, we therefore extend the framework by allowing *frame types as values of attributes*. To this end, we introduce nominals (names, constants) for frame types into the description language, which means to treat frame types as “first class citizens” of the frame models. More formally, we assume that every (open) attribute-value description can give rise to the name of a frame type, which is notationally indicated by enclosing the description in double lines. For example,  $\|\text{causation}\|$  and  $\|\text{broken} \wedge \text{PATIENT} : \text{phys-obj}\|$  are names of frame types. Frame types are related to their instances and to each other by the relations *is-instance-of* (*inst*) and *is-subtype-of* (*subtype*), respectively. For instance,  $\|\text{causation}\|$  *is-subtype-of*  $\|\text{event}\|$  is assumed to be true.

In order to characterize an event with respect to its progression of incremental, ongoing changes, the event is assumed to have an attribute PROG(RESSION) whose value specifies the *type* of the change in question. Processes of drying can then be characterized as having an attribute PROG whose value is the type  $\|\text{becoming-drier}\|$ . More precisely, the type in question is  $\|\text{becoming-drier} \wedge \text{ENTITY } x\|$ , where  $x$  is the entity that is drying. This frame type is to be seen as a shorthand for the more complex



**Fig. 2** Frame-semantic representation of a complex type of incremental-change

type shown in Fig. 2, which provides an explicit decomposition of the underlying change of state: Events of the type in question are events of type *incremental-change* of an ENTITY  $x$  such that the MOISTURE value at the FIN(AL) stage (of  $x$ ) is lower than MOISTURE value at the INITIAL stage.

Characterizing an event  $e$  by  $\text{PROG} \triangleq T$  is meant to express the fact that every (appropriate) *event segment*  $e'$  of  $e$  ( $e' \text{ segm } e$ ) is an instance of the type  $T$  ( $e' \text{ inst } T$ ). That is, the following constraint schema is required to be valid:

$$(12) \quad e \cdot \text{PROG} \triangleq T \wedge e' \text{ segm } e \rightarrow e' \text{ inst } T$$

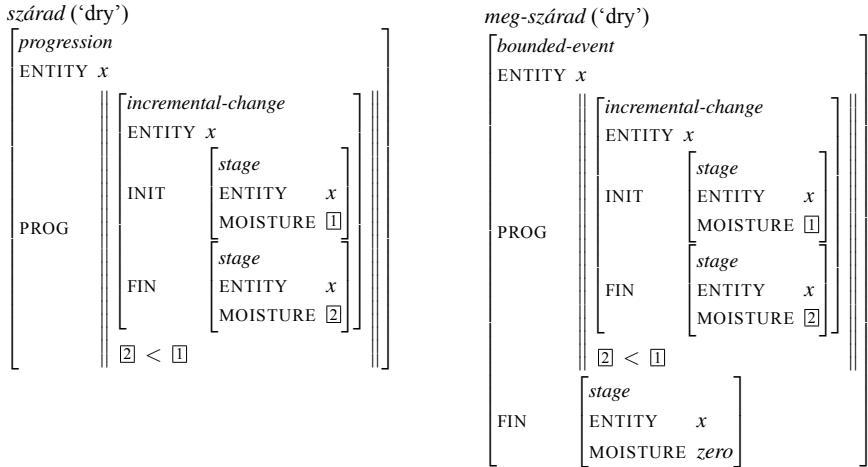
It is this schema that makes explicit the universal quantification over subevents encoded by PROG. Note that (12) applies only to event segments which are referentially introduced. That is, the schema is applied “on demand”.

### 3 Semantic Analysis of Verbal Particles

A central pattern of our analysis is that verbal particles in Hungarian, and other lative-marked verbal modifiers, can turn activity (or process) descriptions into accomplishments by adding a boundary condition to the event frame associated with the verb.<sup>6</sup> Following the outline sketched in Sect. 2, the boundary information is imposed by syntax-driven frame composition on a *scale* or *dimension of change* component within the event.

The frame representation of the drying process and the effect of adding *meg-* is sketched in Fig. 3. The process is modeled as a progression characterized by an

<sup>6</sup>Turning atelic events into telic ones is a rather frequent function of the verbal particle in Hungarian. Note, however, that this function is not always present. As É. Kiss (2008) and Kiefer and Németh (2012) point out, there are particle verbs denoting a static (and hence inherently atelic) event; moreover, duplication of the verbal particle signals iteration (as non habitual repetition), which is atelic as well. In the former group, the base verb is either a perception verb or a verb expressing spatial position. In these cases, the verbal particle contributes directionality.



**Fig. 3** Frame-semantic representation of combining *meg-* with *szárad* ('dry')

incremental decrease of moisture. The value of PROG is the frame type which characterizes the subevents of the progression. The particle *meg-* adds a FINAL attribute to the progression frame, and the constraint shown in (13) picks out the type of the FINAL value from the progression structure.<sup>7</sup>

$$(13) \quad \text{FINAL : stage} \wedge \text{PROG} \parallel \text{FINAL} \parallel \triangleq T \Rightarrow \text{FINAL is-of-type } T$$

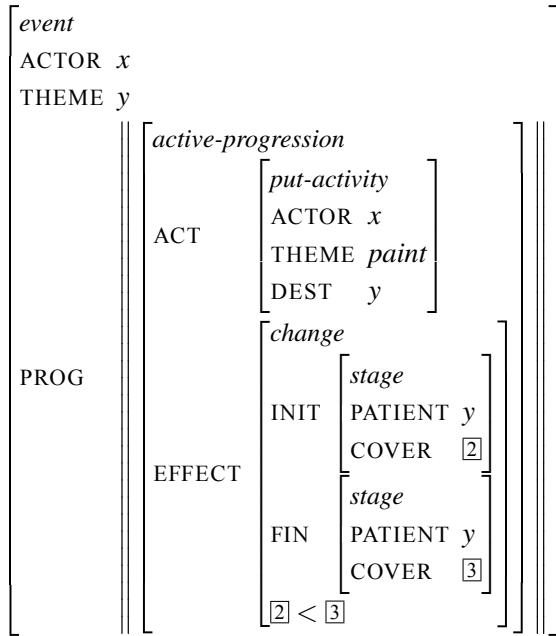
A further constraint enforces the extremal value of the scalar attribute in the final stage.

The above proposal can be directly applied to the analysis of verbs expressing an incremental change. Compare again the sentence without verbal particle in (3b), repeated as (14a), with the sentence in (4b) with the particle *le-* ('down, off') in its default preverbal position, repeated as (14b).

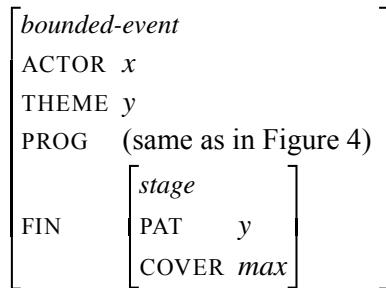
- |                                    |          |
|------------------------------------|----------|
| (14) a. Anna festette a kerítés-t. | [atelic] |
| Anna painted the fence-ACC         |          |
| ‘Anna was painting the fence.’     |          |
| b. Anna le-festette a kerítés-t.   | [telic]  |
| Anna VPTCL-painted the fence-ACC   |          |
| ‘Anna painted the fence.’          |          |

The base verb *fest* ('paint') denotes an event of type *active-progression* which goes along with an incremental change of the theme. More precisely, as indicated by the frame representation shown in Fig. 4, the base verb *fest* expresses an action by the ACTOR  $x$ , affecting the THEME  $y$  by incrementally putting more and more paint on

<sup>7</sup> $\phi \Rightarrow \psi$  is short for  $\forall x(\phi(x) \rightarrow \psi(x))$ , where  $\phi$  and  $\psi$  are one-place predicates.

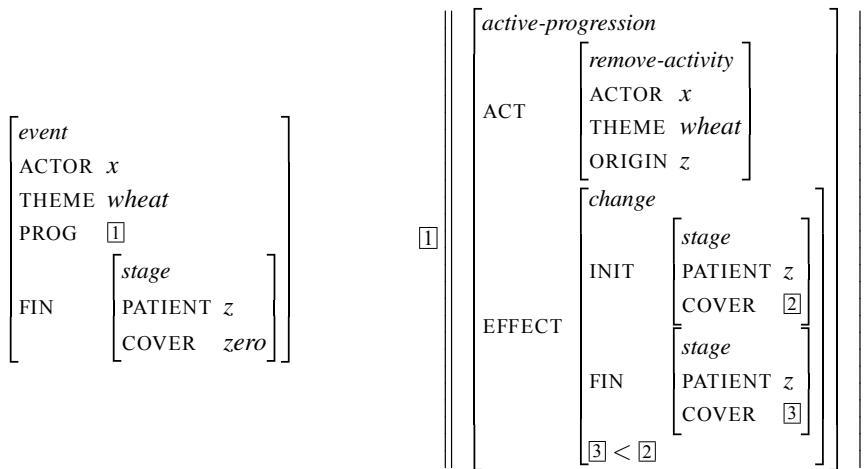


**Fig. 4** Frame representation of *fest* ‘paint’



**Fig. 5** Frame representation of *le-fest* ‘paint’ (bounded)

the surface of *y*. In this incremental change of the surface, for each arbitrary part of the progression it holds that at the final stage of that part the surface is covered more than it was at the initial stage of that given part. In (14b), by comparison, the verbal particle *le-* ('down') contributes the final stage, turning the event into a *bounded event* (Fig. 5).



**Fig. 6** Frame representation of *le-arat* 'reap/harvest'

Consider now example (15) with the particle verb *le-arat* ('reap/harvest').

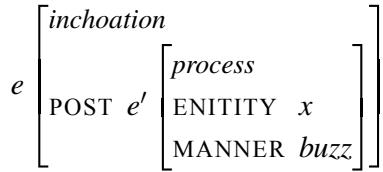
- (15) A gazda le-aratta a búzá-t.  
      the farmer VPTCL-reaped the wheat-ACC  
      'The farmer reaped the wheat.'

The base verb expresses an activity of removing the THEME (wheat) from an unspecified location *z* such that its coverage will incrementally decrease. This is represented in the frame type shown on the right of Fig. 6: the value of the ACT attribute is the *remove-activity* with the ORIGIN *z*, which is identical to the PATIENT of the initial and final stages of the change. Similarly to the previous examples, the verbal particle signals the final stage at which the coverage of *z* is zero (or minimal). Note that in the above example, the THEME of the main event is not identical to the PATIENT of the incremental change. Although less frequent, there are examples of reap/harvest where the changing object is expressed as the direct object of the utterance:

- (16) Le-arat-ták Devecser határ-á-ban péntek-en az első [...]  
      VPTCL-reap-3PL Devecser border-3POSS-INESS Friday-SUPESS the first  
      kísérleti energiaültetvény-t [...]  
      experimental energy.plantation-ACC  
      'At the border of Devecser, the first energy plantation was reaped on Friday [...]

(*Magyar Nemzet Online*, 30 November 2012)

The examples of the inchoative function of the particles *meg-*, *el-* and *fel-* mentioned in Sect. 1 are partially in line with the observations made about the inchoative use of the Russian prefix *za-* as presented by Zinova (2017). The inchoative func-



**Fig. 7** Frame representation of *fel-zúg* ‘begin to buzz’

tion of the verbal particle is compatible with base verbs expressing a *state*, *activity* or *process*. We represent this use as expressing an event of type *inchoation* with a POST(ERIOR) attribute whose value is the posterior event of the inchoation, i.e., a state/activity/process of the type denoted by the base verb; cf. Fig. 7.

## 4 Semantic Composition and the Syntax-Semantics Interface

As to the modeling of the interaction between syntax and semantics we apply the framework of Role and Reference Grammar (RRG; Van Valin and LaPolla 1997; Van Valin 2005). RRG is a surface oriented grammar, developed from a typological perspective and explicitly concerned with the interplay of syntax, semantics and pragmatics. The layered structure of the clause in RRG aims to capture universal characteristics of clause structure in natural languages, while language specific features are expressed via a range of constraints. The layered structure reflects the distinction between predicates, arguments, and non-arguments. The *core* layer consists of the *nucleus*, which specifies the (verbal) predicate, and the syntactic arguments. The *clause* layer contains the core as well as extracted arguments. Each of the layers can have a *periphery* where adjuncts are attached to; cf. Fig. 8 (where ‘RP’ stands for *referential phrase*).

The heart of the grammatical system of RRG is a bi-directional *linking algorithm* between the syntactic and the semantic representations of the sentence, reflecting both processes of production and comprehension. The interaction of syntax and semantics is furthermore influenced by discourse-pragmatics (the focus structure of the utterance) and language-specific constructional schemas. In our analysis we rely on a formalized version of RRG, following Osswald and Kallmeyer (2018), in which tree nodes can carry features. This allows for the elimination of the PRED node, which has no other function than marking the element as predicative, and which can be simply represented by the feature [PRED +]. Features can also be used to establish the link between syntactic elements and the corresponding semantic representations.

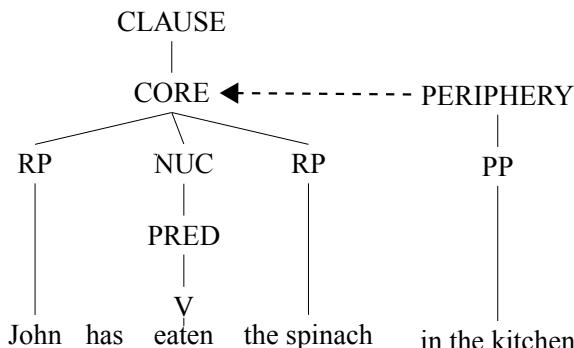
We propose different structural representations for the verbal particle and the resultative predicate. The main difference is that the latter construction is analyzed as a nuclear cosubordination (cf. Van Valin 2005), which corresponds to complex

predicate formation, while the particle is taken as a modifier of the verbal nucleus; cf. Fig. 9. By this distinction we argue against a uniform account of the semantic contribution associated with the preverbal position as proposed by É. Kiss (2008), who claims that the verbal particle in this position functions as a resultative, terminative or locative secondary predicate of the theme argument. This proposal seems to be too restrictive since verbal particles do not necessarily introduce a secondary predicate. Consider, for instance, the particle-verb combination in (17), for which the assumption of a secondary predication is hard to justify.

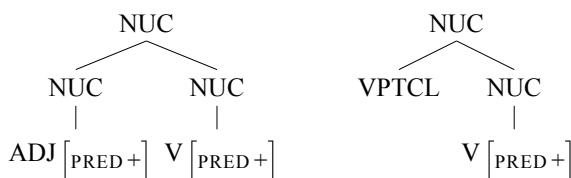
- (17) Anna el-énekt *egy* dal-t.  
 Anna VPTCL-sang a song-ACC  
 ‘Anna has sung a song.’

Based on similar observations, Bene (2009) argues that the verbal particle merely functions as a delimiter rather than as a secondary predicate. In our analysis, we aim to make this distinction explicit by analyzing the construction of a resultative adjective-verb combination as a nuclear cosubordination with two predicative elements and the particle as a modifier of the verbal nucleus.

In the formalized version of RRG introduced in Osswald and Kallmeyer (2018), the syntactic inventory, whose elements are subject to compositional syntactic operations such as substitution and adjunction, consists of *elementary trees* in the sense of Lexicalized Tree Adjoining Grammars (Joshi and Schabes 1997). The elemen-



**Fig. 8** Universal elements in the RRG clause structure

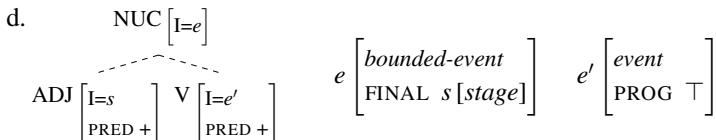
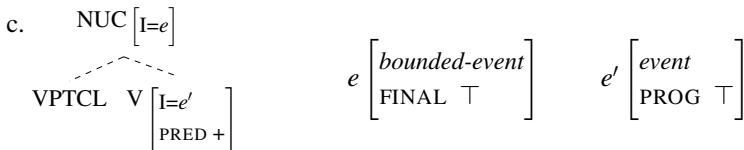
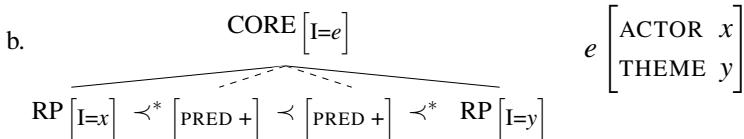
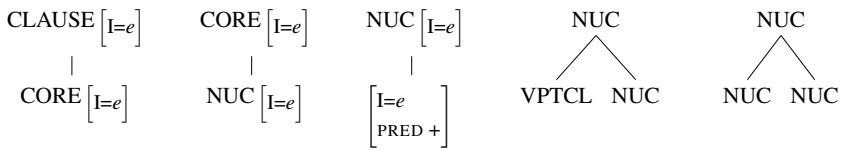


**Fig. 9** Structures for resultative predicates and verbal particles

tary trees encode full argument projections. They are specified in a modular way in the so-called *metagrammar* (Crabbé et al. 2013). The metagrammar is basically a declarative system of tree descriptions about node dominance and precedence which allows one to define classes of grammatically relevant tree constraints. These classes can then be combined to generate the elementary trees as minimal models of the constraints. It is thus the level of the metagrammar where important grammatical generalizations about the elementary constructions of a language are expressed.

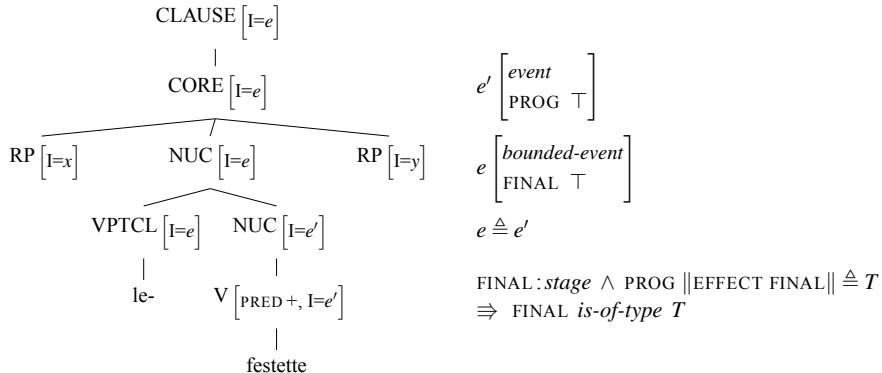
The metagrammar classes used in the analysis of our examples are sketched in (18).<sup>8</sup>

(18) a.



The tree fragment in (18b), together with its semantic contribution, describes a structure with the actor argument in the preverbal field and the theme argument in the postverbal field. The tree fragment in (18c) and its associated semantic contribution describes the verbal particle in its default position and its semantic contribution as adding a boundedness condition to the event. The fragment in (18d) describes a resultative adjective in the preverbal position contributing a final stage  $s$  (boundedness condition) in which a secondary predicate holds.

<sup>8</sup>In the illustrations,  $\prec^*$  stands for precedence,  $\prec$  for immediate precedence, edges by solid lines stand for immediate dominance, and the dashed lines for dominance.



**Fig. 10** Interaction between syntax and semantics for sentence (14b)

Let us apply the proposed analysis to the example in (14c). Figure 10 illustrates the interaction between syntax (in terms of RRG) and frame semantics for particle-verb combinations like *le-festette* ('VPTCL-painted'). The verb *festette* contributes an event  $e'$  with a progression component while the verbal particle contributes a bounded event  $e$  with a final stage. The equation  $e \triangleq e'$  states that these two components both contribute to the same event rather than expressing two separate events. The constraint shown at the lower right of Fig. 10 corresponds to the constraint in (13) and ensures that the final stage of the bounded event and the final stage of the effect of the incremental progression must be of the same type. At the end of the derivation, the semantic composition leads to the representation illustrated in Fig. 5.

As shown in example (1b), repeated as (19), resultative predicates also function as verbal modifiers, occupying the immediate preverbal position.

- (19) Anna zöld-re    festette a    kerítés-t.  
      Anna green-SUBL painted the fence-ACC  
      ‘Anna painted the fence green.’

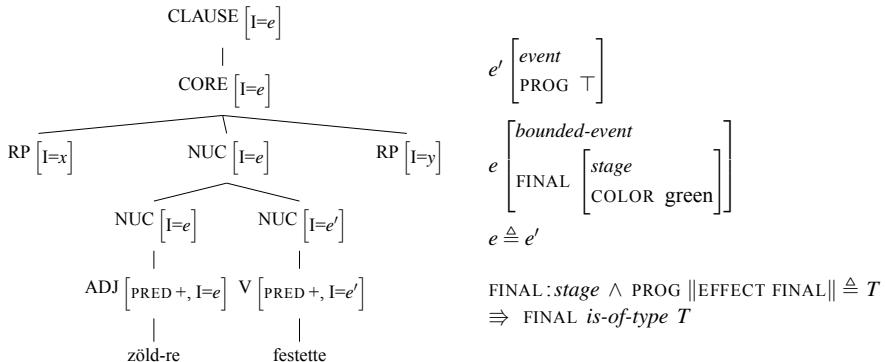
The combination of the preverbal resultative predicate and the verb is analyzed as a nuclear cosubordination with both NUC elements being predicative. The resultative predicate *zöld-re* ('green-SUBL') in its default position also indicates boundedness (telicity), and being predicative it provides a secondary predication of the theme: in the final stage of the changing theme its color is green. The constraint on the final stages is the same as before; cf. Fig. 11.

Verbal particles can also co-occur with resultative predicates, which poses further interesting questions for the syntax-semantics interface. If the particle and the resultative predicate co-occur in a neutral sentence, they cannot both be preverbal. In

this case, the particle occupies the immediate preverbal position while the resultative phrase appears postverbally; see (20a) versus (20b).<sup>9</sup>

- (20) a. Anna le-festette zöld-re a kerítés-t.  
          Anna VPTCL-painted green-SUBL the fence-ACC  
          ‘Anna painted the fence green.’
- b. \*Anna zöld-re le-festette a kerítés-t.  
          Anna green-SUBL VPTCL-painted the fence-ACC

The analysis of (20a) is in line with the analysis of the previous sentences. The particle and the verb form a modified nucleus (*le-festette*) that forms a complex predicate with the resultative predicate (*le-festette zöld-re*) by nuclear cosubordination. The final derivations for the examples in (19) and (20a) lead to the same semantic representation, in accordance with our intuitions; cf. Figs. 11, 12, and 13. Note that the progression component, that is, the representation of the incremental change, is the same in all three cases. Examples (19) and (20a) differ from (14b) in that the latter does not contain a secondary predication but merely a delimiter indicating boundedness (telicity).



**Fig. 11** Interaction between syntax and semantics for example (19)

<sup>9</sup>The linearization in (20b) is grammatical in case the resultative predicate gets a contrastive topic intonation. In this article, we only consider neutral sentences.

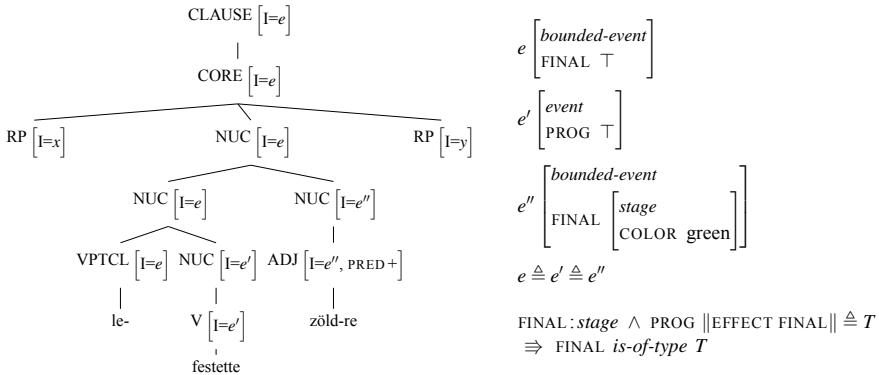


Fig. 12 Interaction between syntax and semantics for example (20a)

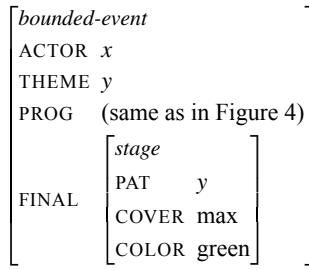


Fig. 13 Derived semantics for (19) and (20a)

## 5 Summary

The main goal of this article was to propose a formal account of the semantic contribution of various verbal particles in Hungarian and to sketch how the semantic representation of the clause can be compositionally derived. We did not aim at a full-fledged descriptive characterization of all the possible particle-verb combinations in Hungarian but concentrated on frequent functions and their formal semantic characterization. While the previous analyses mentioned in Sect. 1 offer adequate insights to the various meaning contributions of the Hungarian verbal particles, they leave open the question of the precise semantic representation and the compositional mechanisms involved. Furthermore, we argued that the characterization of É. Kiss (2008) of the particle as a secondary predication is too strong. Kiefer and Ladányi (2000) and Kiefer (2009) provide a wide coverage descriptive analysis but lack a well-defined formal characterization. They introduce nine productive Aktionsart formations by verbal particles, but without specifying their semantic representation formally. We presented a formal, compositional analysis of some of their basic descriptive insights. We focussed on frequent cases of the telicizing function of verbal particles and sketched a representation of the inchoative meaning contribution.

The use of the framework of decompositional frame semantics proved useful for this purpose as it provides a formal tool for a fine-grained representation of the event structure of the predicate and for the Aktionsart-effects of modified and complex predicates. The semantic characterization of verbal particles in our analysis is close to the analysis of Kardos (2016), among others. The main contribution of our approach is an explicit semantic and syntactic representation and a compositional model.

**Acknowledgements** The research presented in this article was supported by the Collaborative Research Centre 991 “The Structure of Representations in Language, Cognition, and Science” funded by the German Research Foundation (Deutsche Forschungsgemeinschaft). We would like to thank the anonymous reviewers for their valuable comments on earlier drafts of this article.

## References

- Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In A. Lehrer & E. F. Kittay (Eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization* (pp. 21–74). Hillsdale: Lawrence Erlbaum.
- Beavers, J. (2008). Scalar complexity and the structure of events. In J. Dölling, T. Heyde-Zybatow, & M. Schäfer (Eds.), *Event structures in Linguistic form and interpretation, language, context and cognition* (pp. 245–265). Berlin: Mouton de Gruyter.
- Bene, A. (2009). What is really the function of the verbal particle in Hungarian. *Suvremena Lingvistika*, 35(68), 207–220.
- Crabbé, B., Duchier, D., Gardent, C., Le Roux, J., & Parmentier, Y. (2013). XMG: eXtensible MetaGrammar. *Computational Linguistics*, 39(3), 591–629.
- Csirmaz, A. (2012). *Scales, verbs and verbal modifiers*. University of Utah (unpublished manuscript).
- É. Kiss, K. (2008). The function and the syntax of the verbal particle. In K. É. Kiss (Ed.), *Event structure and the left periphery. Studies on Hungarian* (pp. 17–55). Dordrecht: Springer.
- Filip, H. (2008). Events and maximalization: The case of telicity and perfectivity. In S. Rothstein (Ed.), *Theoretical and crosslinguistic approaches to the semantics of aspect* (pp. 217–256). Amsterdam: John Benjamins.
- Forgács, T. (2004). Grammaticalization and preverbs. *Acta Linguistica Hungarica*, 51(1–2), 45–84.
- Hay, J., Kennedy, C., & Levin, B. (1999). Scalar structure underlies telicity in “degree achievements”. In T. Matthews & D. Strolovitch (Eds.), *Proceedings of SALT IX*, Ithaca, NY (pp. 127–144). CLC Publications.
- Joshi, A. K., & Schabes, Y. (1997). Tree-adjoining grammars. In G. Rozenberg & A. Salomaa (Eds.), *Handbook of formal languages. Vol. 3: Beyond words* (pp. 69–123). Berlin: Springer.
- Kagan, O. (2013). Scalarity in the domain of verbal prefixes. *Natural Language and Linguistic Theory*, 31, 483–516.
- Kagan, O. (2016). *Scalarity in the verbal domain: The case of verbal prefixation in Russian*. Cambridge: Cambridge University Press.
- Kallmeyer, L., & Osswald, R. (2013). Syntax-driven semantic frame composition in lexicalized tree adjoining grammars. *Journal of Language Modelling*, 1(2), 267–330.
- Kardos, É. (2012). *Toward a scalar semantic analysis of telicity in Hungarian*. Ph.D. thesis, Debreceni Egyetem.
- Kardos, É. (2016). Telicity marking in Hungarian. *Glossa*, 1(1), 41:1–37.
- Kearns, K. (2007). Telic senses of deadjectival verbs. *Lingua*, 117, 26–66.

- Kennedy, C., & Levin, B. (2008). Measure of change: The adjectival core of degree achievements. In L. McNally & C. Kennedy (Eds.), *Adjectives and adverbs* (pp. 156–182). Oxford: Oxford University Press.
- Kiefer, F. (2009). Types of temporal adverbials and the fine structure of events. In K. É. Kiss (Ed.), *Adverbs and adverbial adjuncts at the interfaces* (pp. 247–269). Berlin/Boston: De Gruyter Mouton.
- Kiefer, F., & Ladányi, M. (2000). Az igekötők [The verbal particles]. In F. Kiefer (Ed.), *Strukturlis magyar nyelvtan, Vol. 3: Morfológia* (pp. 453–518). Budapest: Akadémiai Kiadó.
- Kiefer, F., & Németh, B. (2012). Amikor az igekötő nem telicizál [When the verbal particle does not telicize]. *Nyelvtudományi Közlemények*, 108, 267–273.
- Krifka, M. (1998). The origins of telicity. In S. Rothstein (Ed.), *Events and grammar* (pp. 197–235). Dordrecht: Kluwer Academic Publishers.
- Levin, B., & Rappaport Hovav, M. (2011). Lexical conceptual structure. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics. An international handbook of natural language meaning* (Vol. 1, pp. 420–440). Berlin: Mouton de Gruyter.
- Löbner, S. (2014). Evidence for frames from human language. In T. Gamerschlag, D. Gerland, R. Osswald, & W. Petersen (Eds.), *Frames and concept types* (pp. 23–67). Dordrecht: Springer.
- Osswald, R., & Kallmeyer, L. (2018). Towards a formalization of Role and Reference Grammar. In R. Kailuweit, E. Staudinger, & L. Künkel (Eds.), *Applying and expanding Role and Reference Grammar (NIHIN Studies)* (pp. 355–378). Freiburg: Albert-Ludwigs-Universität, Universitätsbibliothek.
- Osswald, R., & Van Valin, Jr., R. D. (2014). FrameNet, frame structure, and the syntax-semantics interface. In T. Gamerschlag, D. Gerland, R. Osswald, & W. Petersen (Eds.), *Frames and concept types* (pp. 125–156). Dordrecht: Springer.
- Pátvics, P. (2002). Néhány gondolat a magyar igekötők eredetéről, valamint aspektus- és akcímőminőség-jelölő funkciójuk (ki)alakulásáról. *Magyar Nyelvőr*, 126(4), 481–489.
- Rappaport Hovav, M. (2008). Lexicalized meaning and the internal temporal structure of events. In S. Rothstein (Ed.), *Theoretical and crosslinguistic approaches to the semantics of aspect* (pp. 13–42). Amsterdam: John Benjamins.
- Szoltész, K. (1959). *Az ősi magyar igekötők* [The ancient Hungarian verbal particles]. Budapest: Akadémiai Kiadó.
- Van Valin, R. (2005). *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press.
- Van Valin, Jr., R. D., & LaPolla, R. J. (1997). *Syntax*. Cambridge: Cambridge University Press.
- Zinova, J. (2017). *Russian verbal prefixation: A frame semantic analysis*. Ph.D. thesis, Heinrich-Heine-Universität Düsseldorf.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# On the Fictive Reading of German *Steigen* ‘Climb, Rise’: A Frame Account



Thomas Gamerschlag and Wiebke Petersen

**Abstract** Fictive motion, i.e., the figurative stative use of verbs of motion, has attracted much attention in cognitive linguistics as a paradigm case for how basic dynamic concepts are exploited figuratively in concept formation (Langacker 1986; Matsumoto 1996; Talmy 2000; Matlock 2004a, b *inter alia*). In this paper, we present a case study of the fictive motion reading of the German movement verb *steigen* ‘climb, rise’ and explore how it can be related to the various dynamic readings of the verb. In our account of *steigen*, which builds on Gamerschlag, Geuder & Petersen’s (2014) analysis of the dynamic readings of the verb, we contrast the different readings in terms of frames, i.e., recursive attribute-value structures in the sense of Barsalou (1992) and Petersen (2007/2015).

**Keywords** Fictive motion · Verbs of motion · Stative reading of dynamic verbs · *steigen/rise* · Frame analysis

## 1 Introduction

In fictive motion, verbs of motion are applied to describe a stative scenario in which the subject referent usually is a stationary, non-moveable entity. In the most typical cases, the subject refers to some kind of pathlike entity such as a road or a line while the original theme, the moving participant of the literal use of the verb, remains unrealized. A German example of the fictive motion use of some verbs is given in (1) below. As can be seen, fictive motion uses serve to highlight spatial properties of the subject referent: *laufen* ‘run’ combines with the modifier *quer* ‘diagonal’ and a directional PP which specify the location of the scar and its orientation in relation to the cheek. Moreover, *springen* ‘jump’ plus PP identifies the eye as a region where the scar is interrupted, while *landen* ‘land’ locates the final part of the scar in the eyebrow when combined with the PP in (1).

---

T. Gamerschlag (✉) · W. Petersen  
Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany  
e-mail: [gamer@phil.hhu.de](mailto:gamer@phil.hhu.de)

- (1) *Eine [...] Narbe lief quer über seine eine Wange. Sie sprang über sein  
a scar ran diagonally across his one cheek it leapt over his  
Auge und landete in seiner Augenbraue.*<sup>1</sup>  
eye and landed in his eyebrow  
A scar ran diagonally across one of his cheeks, leaping over his eye and landing in one of his eyebrows.'

In German, both manner of motion verbs as well as directed motion verbs allow for fictive readings. In (1), *laufen* ‘run’ and *springen* ‘leap/jump’ are verbs encoding manner, whereas *landen* ‘land’ refers to a downward motion which ends up on some surface. Additional examples of fictive readings of path verbs are given below. For instance, in (2), the verbs *überqueren* ‘cross’ and *abbiegen* ‘turn off (the road)’ are applied to highlight properties of the course of the road.

- (2) *Die Straße überquert den Fluss und biegt dann in Richtung Flughafen ab.*  
the road crosses the river and turns then in direction airport PARTICLE  
'The road crosses the river and then turns in the direction of the airport.'

Likewise, *steigen* ‘rise’, which originally denotes a dynamic change in height of a moveable object, refers to an upward slope of the terrain in (3).

- (3) *Der Weg steigt [...] langsam auf eine Höhe von 4450 m.*<sup>2</sup>  
the trail climbs slowly to a height of 4450 m  
'The trail climbs slowly to a height of 4450 meters.'

The verb *steigen*, variously translated as ‘climb’, ‘rise’ and also ‘step’, is highly polysemous. It exhibits a use as a manner of motion verb in addition to a purely directional reading and an “intensional” (figurative) use, as well as a fictive motion reading as in (3). We consider the meaning of *steigen* as a representative example of a complex array of different verb senses and the way they are systematically interrelated. These different senses will be illustrated in Sect. 3 after a concise overview of previous approaches to fictive motion in Sect. 2. In Sect. 4, we will give a short summary of Gamerschlag et al.’s (2014) frame approach to the dynamic readings of *steigen*. After a closer look at the fictive motion use of the verb in Sect. 5, we will present a frame analysis of this reading in Sect. 6. In Sect. 7 the fictive motion use of *steigen* is compared to the intensional use. Finally, in Sect. 8 we will indicate how the sketch of our frame analysis of fictive motion can be extended and elaborated on in various ways.

---

<sup>1</sup>Example taken from the novel *Der fünfte Spieler* by Blue Balliet, Aufbau Digital 2011.

<sup>2</sup>[www.bhutan-travel.de/index.php/trekking-in-bhutan/mittelschwere-treks/18-trekking-in-bhutan-184-jhomolhari-trek](http://www.bhutan-travel.de/index.php/trekking-in-bhutan/mittelschwere-treks/18-trekking-in-bhutan-184-jhomolhari-trek) (accessed 5 June 2019)

## 2 Previous Accounts of Fictive Motion

Given the confinements of this paper, it is not possible to do justice to all the work that has been done in regard to fictive motion phenomena in the past decades. The recognition of fictive motion as such and its relevance to language, concept formation and cognitive processing is a merit of cognitive linguistics. The term ‘fictive motion’ goes back to work by Leonard Talmy, starting out from the 70’s, developing over the following decades and resulting in insights such as the typology of fictive motion presented in Talmy (2000). Though alternatively referred to by terms such as ‘abstract motion’ (Langacker 1986) and ‘subjective motion’ (Matsumoto 1996), the phenomenon is characterized by a well-defined empirical base which also allows for cross-linguistic comparison. The central claim of cognitive linguists that fictive motion involves the mental simulation of movement or scanning along a path has been corroborated by psycholinguistic research which builds on results from various kinds of experiments. Matlock (2004b) and more recently Matlock and Bergmann (2014) and Hütte and Matlock (2016) give an excellent overview of experimental research on the phenomenon, including their own work. Different kinds of experiments such as narrative understanding tasks and studies based on drawing and eye movement provide evidence that fictive motion goes along with a conceptualizer simulating motion. Matlock (2004b) also shows how assuming mental simulation as part of the concept of fictive motion readings can account for a number of linguistic properties such as the spatial characteristics of the subject referent and the co-occurrence of temporal expressions. In spite of all their insights on the phenomenon, cognitive analyses usually refrain from a formal representation, thereby lacking a level of explicitness necessary for a deeper understanding of fictive motion. Instead, much of the discussion in the cognitive linguistics realm centers around the question of how fictive motion fits into accounts of metaphor and metonymy. For instance, Kövecses (2015) argues against an analysis of fictive motion in terms of conceptual metaphor, since an account of this type would involve an incomplete mapping, leaving components of the dynamic source, such as the moving entity, without a corresponding element in the static target. More recently, stative readings of dynamic verbs have attracted some attention in formal semantics. In his analysis of the stative uses of motion verbs, Gawron (2009) provides an elaborate account of spatial change as opposed to temporal change in which he focuses on so-called “spreading motion” referred to by extent verbs such as *widen* and *cover*. Following Gawron’s ideas, Koontz-Garboden (2010) and Deo et al. (2013) propose accounts of stative uses of dynamic verbs in which the time scale/axis underlying the dynamic use is replaced by a spatial scale/axis. Although these time-to-space transfer analyses elegantly explain a number of properties of stative uses including the co-occurrence of various modifiers, they do not explicitly address fictive motion constructions of the type illustrated above. It is not clear, therefore, how these approaches would account for the range of modifiers that show up as a result of the dynamic origin of fictive motion. In the following sections, we will present a first sketch of a frame analysis of the fictive reading of *steigen* which deals with the range of co-occurring modifiers and the way they are linked to the dynamic source of fictive motion.

### 3 The Four Major Readings of *Steigen* ‘Climb, Rise, Step’

Due to its complex polysemy, the German verb *steigen* ‘climb, rise, step’ is particularly interesting in regard to the question of how the fictive motion use is embedded in the meaning array of a basically dynamic verb of motion. Gamerschlag et al. (2014:116) distinguish the four major uses illustrated in (4).

(4) *steigen*

- a. as a verb of manner of motion

*Die Ziegen steigen auf's Dach / vom Dach (herunter).*  
 the goats climb onto.the roof from.the roof (down)  
 ‘The goats are climbing onto the roof / (down) from the roof.’

- b. as a verb of directed motion

*Der Ballon steigt höher und höher / \*tiefer und tiefer.*  
 the balloon climbs higher and higher deeper and deeper  
 ‘The balloon is climbing higher and higher / \*deeper and deeper.’

- c. as an intensional verb of change along a property scale

*Die Temperatur steigt von 3 auf 10 Grad / \*von 3 auf 10 Grad.*  
 the temperature rises from to degrees from to degrees  
 ‘The temperature is rising from 3 to 10 degrees / \*from 10 to 3 degrees.’

- d. as a static verb of “fictive motion”

*Das Gelände steigt von 750 auf 761 Meter [...]<sup>3</sup> / \*von 761 auf 750 Meter.*  
 the terrain climbs from to meter from to meter  
 ‘The terrain climbs from 750 to 761 meters / \*761 to 750 meters.’

The readings illustrated in (4a) and (b) are literal dynamic uses of the verb which refer to movement in space. They can be differentiated due to a couple of asymmetries. First, *steigen* as a verb of manner of motion (henceforth *steigen<sub>mm</sub>*) requires the use of limbs for the kind of motion referred to. Therefore, only animate subject referents with a suitable anatomy are permitted, such as *Ziegen* ‘goats’ in (4a). It is important to note that *steigen<sub>mm</sub>* is not confined in regard to the direction of motion. As can be seen in (4a), PPs specifying upward as well as downward motion are admissible. Directional *steigen* (henceforth *steigen<sub>dir</sub>*) as in (4b) does not make reference to a particular manner of using one’s limbs. By consequence, the subject referents of *steigen<sub>dir</sub>* can refer to freely suspended entities such as *Ballon* ‘balloon’ in (4b). However, *steigen<sub>dir</sub>* can only denote upward movement as shown by the non-admissibility of a modifier specifying a downward path. This asymmetry in regard to admissible directional complements correlates with their omissibility: While directional PPs can be left out with *steigen<sub>dir</sub>* they cannot be omitted with *steigen<sub>mm</sub>*.

The example in (4c) illustrates a figurative use of *steigen* which abstracts away from spatial motion while referring to abstract “motion” along a scale, such as

---

<sup>3</sup><https://www.suedkurier.de/region/bodenseekreis-ober schwaben/heiligenberg/Neues-Wohnen-und-Arbeiten-in-Heiligenberg;art372476,8460587> (accessed 5 June 2019)

the temperature scale introduced by the subject. Following the formal analyses by Montague (1973) and Löbner (1979, 1981) among others, we will refer to this use as ‘intensional *steigen*’ (henceforth *steigen<sub>ins</sub>*). Characteristically, this use involves a total change of the subject referent over time, as opposed to the partial change of the subject referent in the literal readings in the first two examples in (4), in which the subject referent only changes with respect to a single dimension, namely its spatial location. Like *steigen<sub>dir</sub>*, *steigen<sub>ins</sub>* can only express an increase along the respective scale but never a decrease. In spite of its abstractness, *steigen<sub>ins</sub>* refers to a true dynamic change within a particular value space. By consequence, it can be grouped together with the two literal meanings given in (4).

In contrast, the fictive motion use of *steigen* (henceforth *steigen<sub>fict</sub>*) does not involve motion interpreted as a dynamic change during the course of the event. Instead, it refers to a stative spatial scenario in which the subject referent is a stationary, usually not moveable entity characterized as having some gain in height. For instance, in (4d) it is specified that the slope of the referent of *Gelände* ‘terrain’ has a positive difference in height of 500 meters between some non-realized starting and end point. As with *steigen<sub>dir</sub>* and *steigen<sub>ins</sub>*, *steigen<sub>fict</sub>* (i) allows for an absolute use and (ii) can only refer to upward ‘fictive’ motion while downward motion is excluded, as shown by the non-admissibility of a negative height difference. In this regard *steigen* parallels English *climb* whose fictive motion use is also restricted to a positive difference in height, thereby relating it more closely to the dynamic directional use of *climb* while setting it apart from the manner reading (cf. Fillmore 1982; Jackendoff 1985; Matsumoto 1996).

Note that many speakers seem to have some preference to use *steigen* in its fictive use with a verbal particle such as *an* ‘up(wards)’ rather than choosing the particleless variant, which is often judged as less felicitous or incomplete. However, the argument that *steigen<sub>fict</sub>* is restricted to a positive gain in height can only be made on the base of the particleless variant since in the case of the complex verb *ansteigen* ‘ascend, move upwards’ one may argue that the upward direction is solely contributed by the particle, while the verb itself could be analyzed as being indifferent with regard to the direction of the path. Likewise, the frame account proposed by Gamerschlag et al. (2014) covers only the (non-fictive) simplex uses of *steigen*. Since our analysis of *steigen<sub>fict</sub>* directly builds on their approach, we will focus on the fictive use of *steigen* without the particle. Nonetheless, a complete understanding of *steigen<sub>fict</sub>* requires a discussion of its relation to the fictive readings of *steigen* plus particle which, however, is beyond the limits of this paper.<sup>4</sup> In order to not rely solely on introspection, we have drawn the examples of *steigen<sub>fict</sub>* mainly from internet sources, being well aware of the unreliability of data of this sort.

In the following sections, we will propose an analysis of *steigen<sub>fict</sub>* in which its meaning is derived from that of *steigen<sub>dir</sub>*, due to similar semantic restrictions.

---

<sup>4</sup>The need for analyzing *steigen<sub>fict</sub>* in relation to the fictive uses of corresponding particle verbs such as *ansteigen* and *aufsteigen* ‘ascend/move upwards’ was pointed out to us by one of the reviewers. The same reviewer also stated that according to his/her grammaticality judgements the fictive use of simplex *steigen* is in principle unproblematic.

Starting from the frame representations of the two literal uses in (4), we will show that the frames of both *steigen*<sub>fict</sub> as well as *steigen*<sub>ins</sub> result from structural operations on the frame of *steigen*<sub>dir</sub> which are necessary to accommodate the frame of the subject referent. Before going into the details of our analysis, we will first give a short introduction into the frame model we adopt.

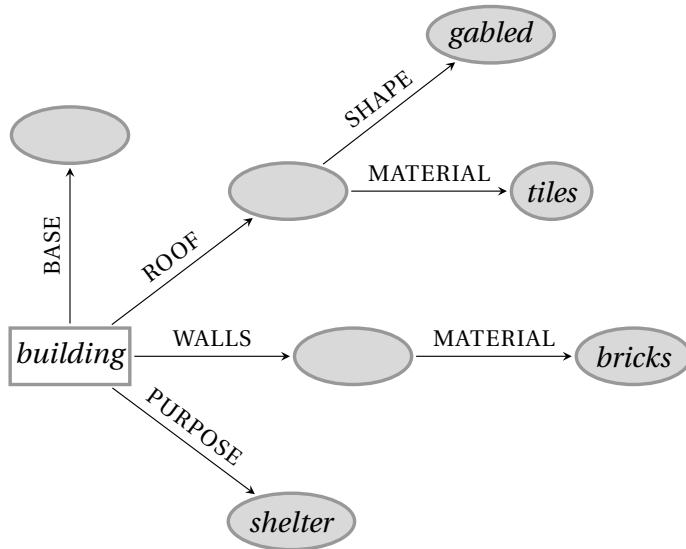
## 4 Frame Analysis of Dynamic *Steigen*: Manner and Directional Reading

### 4.1 *Frames for Objects*

The participants of an event denoted by a verb can be many different kinds of different objects. Usually, these objects are the referents of nominal concepts introduced by noun phrases. Following Barsalou's (1992) idea that conceptual knowledge is represented by means of frames, which provide an explicit, variable-free, and cognitively plausible representation format, we assume that nominal concepts are best captured by frame representations. More precisely, we build on Löbner's (2011) theory of nominal concept types and Petersen's (2007/2015) formalization of frames according to which frames are defined as recursive attribute-value structures with the attributes corresponding to mathematical functions. For illustration, the graph representation of the object concept 'building with brick walls and gabled tiled roof' is given in Fig. 1 below.

The central node specifies the referent of the frame, in this case a particular type of building. The referent is characterized by the attributes branching off the central node: The mereological attributes ROOF, WALLS, and BASE map the referent to particular parts of it. In addition, the value of the attribute PURPOSE points to the function of the building to serve as some kind of shelter. Frames are characterized by their recursive potential, allowing for zooming into the nodes by expanding them into additional attribute-value pairs. For instance, the value of ROOF has the two attributes SHAPE and MATERIAL, each of which comes with particular values. Note that the frame graph in Fig. 1 is kept reasonably simple for the sake of illustration. In principle, frame representations can be unlimitedly detailed by specifying additional attributes and their possibly complex values.

In spite of their flexibility, the range of frames is not arbitrary in the model we adopt. Rather, frames are determined by a type signature that specifies admissible attributes and the type of values they can take. Type signatures model conceptual knowledge and express all kinds of learned constraints such as hierarchical relations, the set of attributes which are adequate for frames of a given type, as well as value restrictions and value dependencies (cf. Petersen 2007/2015 for details).



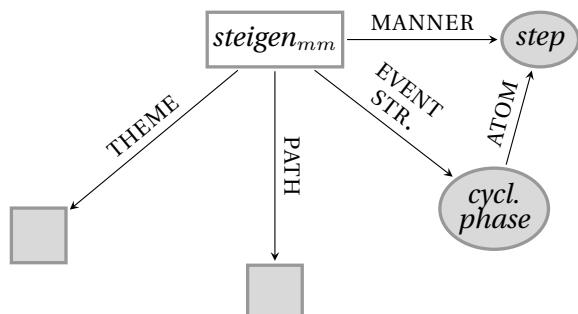
**Fig. 1** Frame representation of ‘building with brick walls and gabled tiled roof’

#### 4.2 *Steigen<sub>mm</sub>*

When it comes to the frames of verbs, things get more complicated since time and change come into play. Following Naumann’s (2013) model of verb frames, a verbal concept can be represented by an overall event frame which represents the global properties of this event. This frame is static in the sense that it does not change during the event. Gamerschlag et al. (2014) assume the static event frame (SEF) for *steigen<sub>mm</sub>* in Fig. 2 below.

The frame representation in the figure above expresses the relations of the objects involved in an event of that sort: *steigen<sub>mm</sub>* has a theme and a path argument which are satisfied by syntactic complements. In the representation, this is indicated by open argument slots marked by square nodes. Moreover, *steigen<sub>mm</sub>* is executed in

**Fig. 2** Static event frame of *steigen<sub>mm</sub>*



a particular manner characterized by step(s) which are the atoms of its internally cyclic event structure. Note that although a typical *steigen<sub>mm</sub>*-event consists of a continuous repetition of steps, it can also be instantiated by a single step, as pointed out by Geuder and Weisgerber (2008).

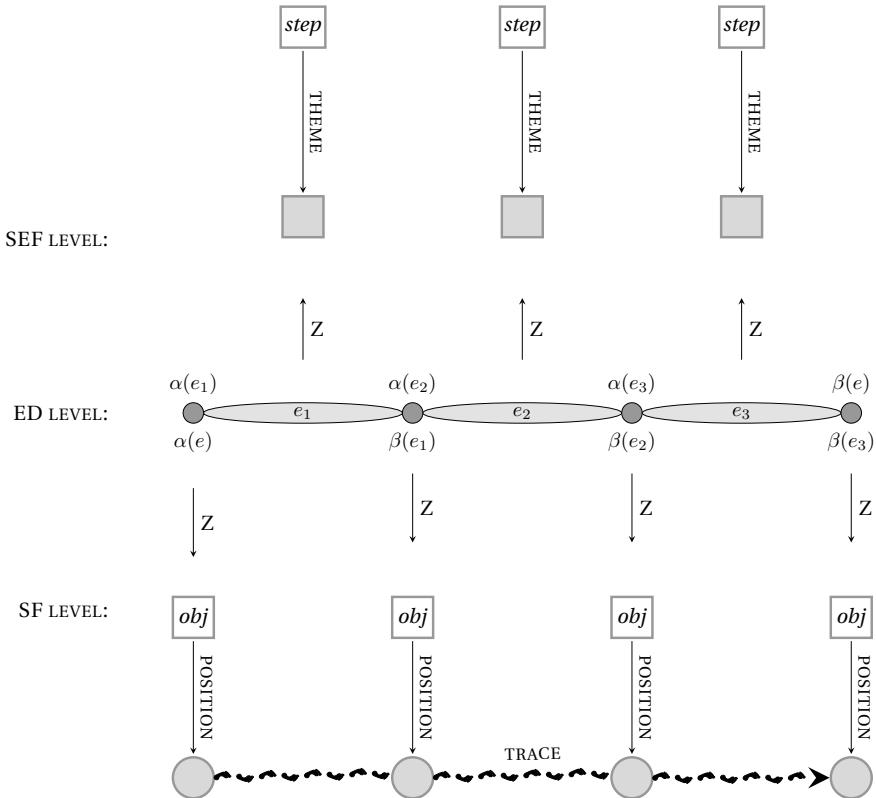
The static event frame is not satisfactory as the sole representation of a dynamic event denoted by *steigen<sub>mm</sub>*. In order to temporalize frames, they need to be related more explicitly to event structure. To this end, Naumann (2013) assumes a three-level model of event representation, which can only be sketched here for reasons of space (see Naumann 2013 and Gamerschlag et al. 2014 for details).<sup>5</sup> First, in addition to the level of static event frames, a level of event decomposition (ED) is required which refers to the temporal structure of an event. In the case of *steigen<sub>mm</sub>*, event decomposition results in a sequence of atomic *step*-subevents e1, e2, ... as shown in the middle of Fig. 3. These subevents are linked to the relevant parts of the static event frame by a zoom function Z such that each atom consists of a single step executed by the theme. As a third level, the situation frame-level (SF) at the bottom of Fig. 3 captures the event-related changes of the participants during the course of the event. In the case of an event structure consisting of atoms, the SF-level provides snapshots of the entity's state at the boundary of each atom. For *steigen<sub>mm</sub>* this means that the change of position of the moving entity (i.e., the subject referent) after each step is specified at this level. Again, the zoom function works as a linking device between the two levels by mapping boundary events to situation frames.

Given the model introduced above, Gamerschlag et al. (2014) assume the frame of *steigen<sub>mm</sub>* in Fig. 4, which results from expanding the manner component into a detailed subframe. This subframe provides information on the force constellation involved by characterizing it as a noticeable, upwards-directed force that is exerted by legs against a solid antagonist.

Note that the frame in the figure above is not static since it reflects the changing location of the subject referent captured at the SF level in Fig. 3. Rather, this frame is some kind of condensed representation that also contains dynamic aspects of the three-level representation outlined above. This is achieved technically by the dynamic attribute TRACE which links the POSITION of the THEME of *steigen<sub>mm</sub>* to its PATH specification. More precisely, TRACE is an attribute that is projected into this frame from the event decomposition frame and maps the changing POSITION of the THEME value to the record of its trace in the time span of the event. Because of their special status, dynamic attributes are indicated by broken lines in the frame graphs.

---

<sup>5</sup>Löbner (2017) proposes an alternative account for capturing change of state verbs in terms of Barsalou frames using first-order comparators. Due to lack of space we cannot discuss his approach and how it can be adopted for the analysis of fictive motion by mapping a change in time onto a change in space.

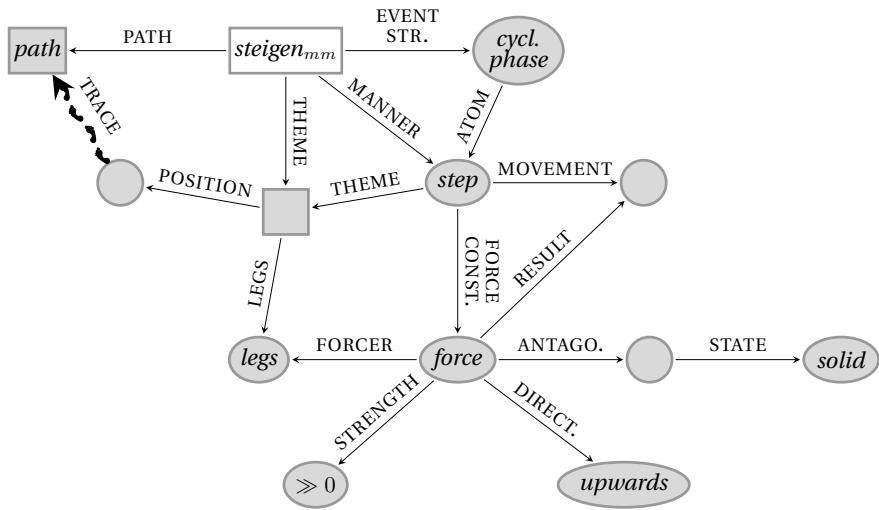


**Fig. 3** Event structure of *steigen<sub>mm</sub>*

#### 4.3 *Steigen<sub>dir</sub>*

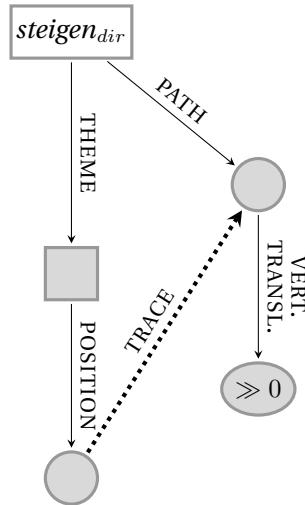
As outlined in Sect. 3, *steigen<sub>dir</sub>* differs from *steigen<sub>mm</sub>* in that it refers to the movement of a freely suspended object without requiring the use of limbs. At the same time, *steigen<sub>dir</sub>* is more restricted than *steigen<sub>mm</sub>* since it can only refer to upward movement. Figure 5 shows the condensed event frame of *steigen<sub>dir</sub>*.

As can be seen, the rich manner component of the frame of *steigen<sub>mm</sub>* is not present in the frame of *steigen<sub>dir</sub>*. As a consequence, the selectional restrictions of *steigen<sub>mm</sub>* do not hold for *steigen<sub>dir</sub>*. Moreover, due to the absence of the step-atoms of the manner component, the event structure is not cyclic anymore but can rather be characterized as a continuous phase. As a further contrast to the frame for *steigen<sub>mm</sub>*, the values of PATH are confined to expressing upward movement. However, apart from the value restriction of the PATH-attribute, the frame component referring to the theme's changing position and the formation of the path by means of the TRACE-function is shared by the condensed frames of both readings. In our analysis, we will show how the frame of *steigen<sub>fict</sub>* can be derived from the frame of *steigen<sub>dir</sub>*.



**Fig. 4** Condensed event frame of *steigen<sub>mm</sub>*

**Fig. 5** Condensed event frame of *steigen<sub>dir</sub>*



## 5 *Steigen<sub>fict</sub>*: Admissible Modifiers and Subject Referents

Before outlining our account of *steigen<sub>fict</sub>* in Sect. 6, we will first have a short look at the range of admissible modifiers and subject referents found with this reading. In addition to permitting adverbial modifiers referring to upward motion, *steigen<sub>fict</sub>*

can combine with adverbs specifying properties such as the slope and the shape of a path, as shown by the examples in (5) and (6).

- (5) *Der Pfad steigt steil / sanft auf den Gipfel.*  
 the trail climbs steeply / gently to the summit.  
 'The trail climbs steeply/gently to the summit.'
- (6) *Die neu asphaltierte Straße [...] steigt kurvenreich auf eine Art Hochplateau.*<sup>6</sup>  
 the newly asphalted road climbs in.serpentines to a kind.of plateau  
 'The newly asphalted road winds upwards (lit.: climbs in serpentines) to some kind of plateau.'

Moreover, adverbs such as *schnell* 'quickly' and *langsam* 'slowly' which are normally associated with temporal properties of dynamic concepts naturally occur with the fictive use, as shown in (7) below. In addition, even modifiers such as *mühsam* 'strenuously' and *gemütlich* 'comfortably', which specify the way a human mover would experience real motion, are admissible.

- (7) *Der Weg steigt schnell / langsam / mühsam / gemütlich auf den Gipfel.*  
 the trail climbs quickly / slowly / strenuously / comfortably to the summit  
 'The trail climbs quickly / slowly / strenuously /comfortably to the summit.'

Another aspect relevant for the understanding of the fictive motion use is the range of admissible subject referents illustrated by the examples in (8). As can be seen, subject referents are not confined to traversable entities such as 'way' and 'road' in German: In (8a) and (b) the referents of *Arteria* 'artery' and *Rohr* 'pipe' are not traversable by humans. However, they still qualify as pathlike entities accessible for mental scanning. Moreover, in German the subject referents need not even be pathlike, as illustrated by (4d) in which a subject such as *Gelände* 'terrain' refers to a two-dimensional space. In our analysis, we will argue that subject referents of this type are licensed because they can be conceived of as embedding the path along which fictive motion can proceed. Likewise, the subject *Wald* 'forest' in (8c) can be interpreted as a two-dimensional entity referring to a specific area or region. Moreover, as shown by the examples in (8d) and (e), even subjects denoting three-dimensional entities are admissible if they provide prominent object sides that restrain possible paths of fictive motion. In these examples it is the (vertical) surface of the mountains and the skyscraper which contains the relevant path. Note that three-dimensional objects of the type illustrated in (8d) and (e) need to have a prominent vertical axis and a

---

<sup>6</sup><http://doczz.net/doc/301001--hilti-foundation> (accessed 5 June 2019)

considerable height ruling out e.g. small objects such as bottles and candles which prototypically have a prominent vertical but are only of small height.<sup>7</sup>

- (8) a. *Die Arteria carotis externa steigt senkrecht nach oben.*<sup>8</sup>  
the arteria carotis externa rises vertically upwards  
'The arteria carotis externa rises vertically upwards.'
- b. *Das Rohr steigt senkrecht durch das Dach.*<sup>9</sup>  
the pipe rises vertically through the roof  
'The pipe rises vertically through the roof.'
- c. *Der Wald steigt [...] bis auf 1'870 m ü[ber] M[eeresspiegel]*<sup>10</sup>  
the forest rises up to 1,870 m above sea level  
'The forest rises to 1,870 meters above sea level.'
- d. *Das Gebirge steigt in unmittelbarer Nähe der Küste [...].*  
the mountains rises in immediate proximity of the coast  
*auf 4000 Höhenmeter.*<sup>11</sup>  
to 4000 meters.in.height  
'The mountains rise up to 4000 meters in height close to the coast.'
- e. *Das Hochhaus steigt siebzig Meter in die Höhe [...].*<sup>12</sup>  
the skyscraper rises 70 meters upwards  
'The skyscraper rises 70 meters into the air.'

As already pointed out by Matsumoto (1996), the availability of non-traversable subject referents is language-dependent. For instance, while English and German are fairly liberal with respect to non-traversable subject referents, according to Matsumoto Japanese is more restricted, excluding subjects referring to walls and fences while allowing for wires and borders to appear as subject referents in fictive motion constructions. However, as observed by Matlock (2004a), even languages such as English and German are sensitive to the property of being traversable.

<sup>7</sup>One reviewer points out that s/he cannot accept three-dimensional subject referents with *steigenfict* while subjects denoting some kind of path or plane are fine. We agree with the reviewer that subject referents of the latter nature are prototypically found with this reading whereas subjects denoting entities of the former kind are more at the periphery of this use and may also vary with respect to native speakers' judgements. However, instances of *steigenfict* plus three-dimensional subject referents, whose grammaticality is also in line with our own judgements, need to be taken into account in a full-fledged analysis of this reading of *steigen*. Due to the lack of space and empirical data, we present some tentative frame account of this subtype of *steigenfict* in Sect. 6 but will refrain from elaborating on it apart from this sketch.

<sup>8</sup>Example taken from I. Bergstrand et al. (eds.) 1964. *Röntgendiagnostik des Herzens und der Gefäße*, p. 655. Berlin: Springer.

<sup>9</sup>Example taken from *Allgemeine medizinische Zeitung mit Berücksichtigung des Neuesten und Interessantesten der allgemeinen Naturkunde*, issue of year 1835, p. 1507. Brockhaus.

<sup>10</sup>[https://www.ur.ch/\\_docn/35377/22.pdf](https://www.ur.ch/_docn/35377/22.pdf) (accessed 5 June 2019)

<sup>11</sup><https://zentralafrika.de/Nationalparks/Mount-Kamerun/> (accessed 5 June 2019)

<sup>12</sup>Example taken from *Hochparterre: Zeitschrift für Architektur und Design*, vol. 27 (2014), p. 14.

According to Matlock (2004a:231f) only “paths ordinarily associated with motion” allow for “information about the way the mover moved, for instance, quickly, slowly, erratically, effortfully [...].” Matlock (2004a:231f) illustrates this observation with the following contrast.

- (9) a. *The highway crawls through the city.*  
 b. ??*The underground cable crawls from Capitola to Aptos.*

The construction in (9a) is felicitous because the subject refers to an entity which was constructed precisely for traveling and therefore is compatible with the particular manner of motion expressed by *crawl*, i.e. progressing slowly and laboriously. In contrast, the example in (9b) is ruled out because a human experiencer cannot be conceptualized as moving on an underground cable in this manner. Likewise, the use of *climb* as a translation of *steigen*<sub>fict</sub> is only felicitous in cases of traversable subject referents since climbing implies the use of hands/feet whereas *rise*, which does not contain manner information of this kind, can be applied in combination with non-traversable subject referents.

Matlock’s constraint is not confined to manner information expressed by the verb. Analogously, some external modifiers yield awkward results if they co-occur with subjects associated with non-traversable paths. As shown in (8b) (repeated as (10)) a non-traversable subject referent such as *Rohr* ‘pipe’ allows for modifiers such as *senkrecht* ‘vertical’ which specify the slope of the path. However, modifiers such as *schnell* ‘quickly’ and *mühsam* ‘strenuously’, which relate to a human moving along a travellable path, are excluded.

- (10) *Das Rohr steigt senkrecht/??schnell / ??mühsam durch das Dach.*  
 the pipe rises vertically/quickly/strenuously through the roof  
 ‘The pipe rises vertically /??quickly/??strenuously through the roof.’

Obviously, the awkward combinations in (10) are ruled out because of some kind of clash between a non-traversable path denoted by the subject and the concept of a human moving along a path suitable for motion evoked by the context.

Given the range of modifiers and subject referents in the examples above, it becomes evident that a proper treatment of instances of fictive motion requires detailed access to properties of the subject referent. In the following section, we will show that the flexibility of frame representations allows for explicit reference to the relevant properties. In particular, we will address the contrastive array of admissible modifiers in dependence of the travellable/non-traversable distinction.

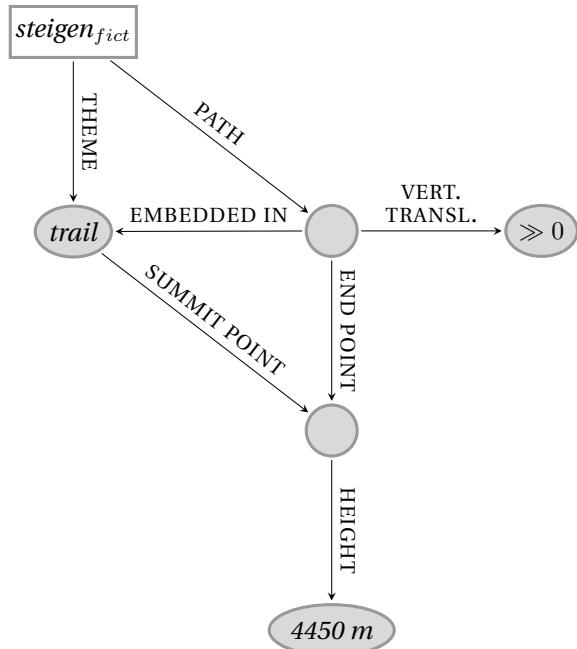
## 6 Frame Analysis of *Steigen*<sub>fict</sub>

For an approach to the fictive reading of *steigen*, we begin with the example in (11), which is a simplified version of the sentence in (3).

- (11) *Der Weg steigt [...] auf eine Höhe von 4450 m.*  
 the trail climbs to a height of 4450 m  
 ‘The trail climbs to a height of 4450 meters.’

Given the fact that *steigen*<sub>fict</sub> is restricted to upward ‘movement’ just as *steigen*<sub>dir</sub>, it is plausible to assume that the meaning of *steigen*<sub>fict</sub> is more closely related to *steigen*<sub>dir</sub> than to *steigen*<sub>mm</sub>. Starting from this observation, our idea goes as follows: If the subject refers to a stationary, non-moveable entity, the literal interpretations of *steigen* are both blocked due to a violation of sortal restrictions with respect to the subject referent. However, in spite of this, the subject referent of *steigen*<sub>fict</sub> can be accommodated by associating it with some suitable part of the existing frames of the literal readings of *steigen*. The value of the PATH-attribute in the frames for both of the literal readings is an entity that can be conceptualized as being embedded in the referent of the subject of *steigen*<sub>fict</sub>. In this regard, both literal readings are appropriate for incorporating the stationary subject referent. However, the frame of *steigen*<sub>dir</sub> is more suited to accommodate the new subject referent since it (a) is more explicit by specifying a path with an upward direction and (b) involves a minor loss of original meaning compared to *steigen*<sub>mm</sub>, which would go along with the deactivation of manner information when combined with a non-appropriate stationary subject referent. Based on these considerations, we assume the frame in Fig. 6 as a representation of the example given in (11) above:

**Fig. 6** Frame representation of *Der Weg steigt auf eine Höhe von 4450 m.* ‘The trail climbs to a height of 4450 m.’

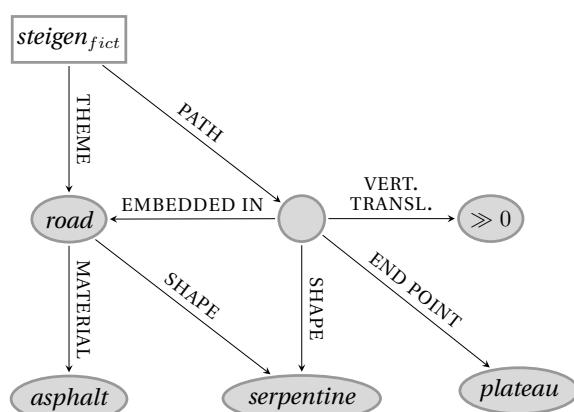


This frame is derived from that of *steigen<sub>dir</sub>* in the following way: First, the stationary subject referent is accommodated in the frame as a new THEME in which the path is embedded. A THEME suitable for that is, for instance, pathlike itself or exhibits a prominent surface that can accommodate a rising path. Second, the original THEME (i.e. the mover) is blocked from being realized which results in deactivation of the meaning components related to actual movement and, consequently, in arriving at the stativized interpretation characteristic of *steigen<sub>fict</sub>*. Due to the value restriction inherited by *steigen<sub>dir</sub>*, the value of VERTICAL TRANSLATION is restricted to a positive value. By consequence, the path can only be conceptualized as having an upward orientation. In addition, spatial modifiers such as *auf 4450 m Höhe* 'to a height of 4450 meters' further restrict the path value by activating additional attributes such as HEIGHT of ENDPOINT. Note that the value of ENDPOINT is shared with the attribute SUMMIT POINT of the theme. By consequence, the HEIGHT of the SUMMIT POINT is identified with the HEIGHT of the ENDPOINT of the path. Furthermore, it is important to note that the frame thus specifies a property of the theme, which is at the same time restricted by a property of the path. Next consider the example in (12), which is a simplification of the one given in (6).

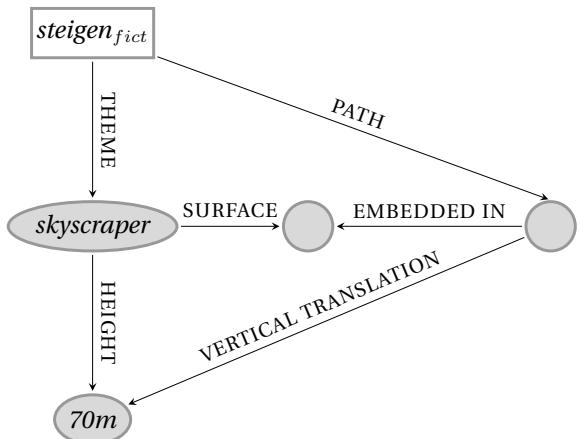
- (12) *Die asphaltierte Straße steigt kurvenreich auf ein Hochplateau.*  
 the asphalted road climbs in.serpentines to a plateau  
 'The asphalted road winds upwards (lit.: climbs in serpentines) to a plateau.'

As shown in the representation of the sentence in Fig. 7, the modifier *kurvenreich* 'winding/in serpentines' evokes the PATH attribute SHAPE for which it highlights a particular value. This attribute is a direct attribute of the path object but its value is again shared with the SHAPE attribute of the theme. As in the preceding example, this ensures that some property of the theme is specified by the construction. As a general rule, we assume that an adverbial modifier of *steigen<sub>fict</sub>* is admissible if it

**Fig. 7** Frame representation of *Die asphaltierte Straße steigt kurvenreich auf ein Hochplateau*. 'The asphalted road winds upwards (lit.: climbs in serpentines) to a plateau'



**Fig. 8** Frame representation of *Das Hochhaus steigt siebzig Meter in die Höhe.*  
 ‘The skyscraper rises 70 meters into the air.’



explicates a value of an attribute of the theme that is restricted by some property of the path.<sup>13</sup>

The example repeated in (13) exhibits a non-pathlike, three-dimensional subject referent.

- (13) *Das Hochhaus steigt siebzig Meter in die Höhe [...].*  
 the skyscraper rises 70 meters upwards  
 ‘The skyscraper rises 70 meters into the air.’

Again, as shown in Fig. 8, the subject referent is accommodated in the frame via the EMBEDDED IN-attribute. More precisely, for three-dimensional entities such as a skyscraper, we assume that the path is embedded in their SURFACE, since it is this part which is accessible for visual scanning. In addition, (13) is interpreted in such a way that the VERTICAL TRANSLATION of the path and the HEIGHT of the skyscraper share the same value.

The use of *steigen<sub>fict</sub>* with non-pathlike subject referents of the type illustrated above appears to be highly restricted, requiring entities with a long and very prominent vertical axis. A better understanding of this combination requires further research that goes beyond the scope of this paper. Therefore, we consider the representation given in Fig. 8 to be only a first approximation of an analysis.

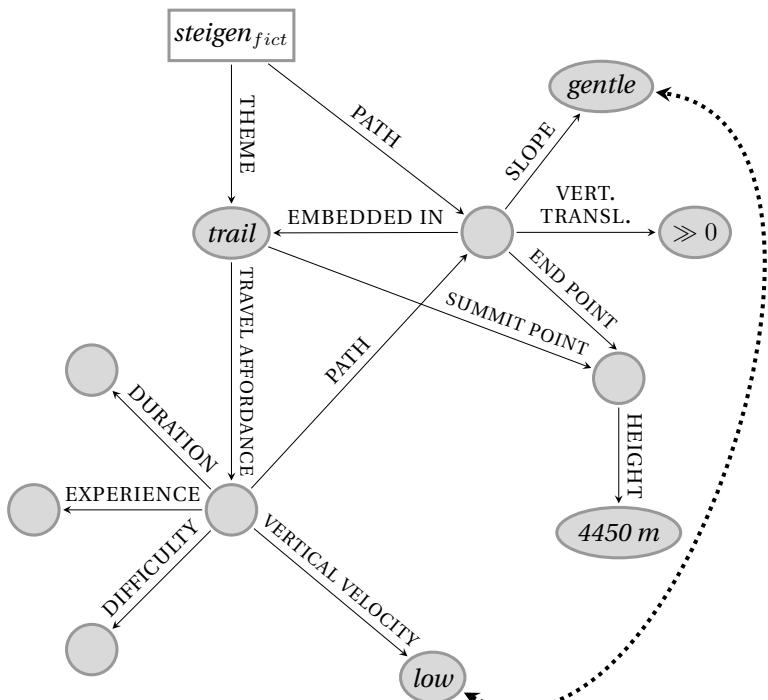
So far, the constraint that the adverbial modifier has to be restricted by some property of the path could be captured in the frame representation by means of value sharing between an attribute of the path and an attribute of the theme. However, if one considers the whole array of admissible modifiers such as the adverb *langsam* ‘slowly’ in (3) repeated in (14) below, it becomes evident that not each instance of *steigen<sub>fict</sub>* can be dealt with in this way.

<sup>13</sup>Similar restrictions on fictive motion expressions have already been proposed by Matsumoto (1996:194).

- (14) *Der Weg steigt [...] langsam auf eine Höhe von 4450 m.*  
 the trail climbs slowly to a height of 4450 m  
 ‘The trail climbs slowly to a height of 4450 meters.’

As argued above, modifiers related to real motion are only licensed if the subject referent provides a traversable path. We assume that this crucial property can be captured by means of an AFFORDANCE attribute understood in the original sense coined by Gibson (1977, 1979) as denoting “action possibilities provided to the actor by the environment (Kaptelinin 2013).” In the case of a subject referent suited for human travel we refer to the relevant attribute as TRAVEL AFFORDANCE as shown in Fig. 9. The value of TRAVEL AFFORDANCE is complex and licenses travel-related attributes such as VELOCITY, DURATION, DIFFICULTY, and EXPERIENCE. Moreover, it exhibits a PATH-attribute which shares its value with the PATH-attribute of the root-node. By consequence, the value of TRAVEL AFFORDANCE varies depending on the particular instantiation of the value of PATH.

As mentioned earlier in this paper, experimental research has convincingly shown that the fictive motion uses of verbs come along with some kind of simulation of actual motion. Since the AFFORDANCE component is a representation of “action possibilities” associated with *steigen*<sub>fict</sub>, it can be regarded as a direct reflex of this



**Fig. 9** Frame representation of *Der Weg steigt langsam auf eine Höhe von 4450 m.* ‘The trail climbs slowly to a height of 4450 m’

kind of simulation with the value of the PATH attribute of TRAVEL AFFORDANCE corresponding to the path that comes about as a result of mental scanning.

For a temporal modifier such as *langsam* ‘slowly’ in (14), we assume that it can be integrated into the frame representation as part of the affordance component as a low value of the attribute VERTICAL VELOCITY, which refers to the speed with which the height of a mover changes. This attribute-value pair is typically correlated with a gentle SLOPE, which is an attribute of PATH.

This correlation between the values of VERTICAL VELOCITY and SLOPE is given only for some average travel velocity of the mover which is contextually specified. Of course, one can also think of a high VERTICAL VELOCITY and a gentle SLOPE or a low VERTICAL VELOCITY and a steep SLOPE. However, this presupposes travel velocities above or below some contextually specified standard for travel velocity.<sup>14</sup>

As a general rule for the admissibility of a modifier of *steigen<sub>fict</sub>* in terms of frames, we assume the following.

- (15) A modifier of *steigen<sub>fict</sub>* is admissible iff it restricts the value of the PATH attribute by either specifying a value of an attribute of the PATH node which is shared with an attribute of the THEME node or by specifying the value of an attribute of the TRAVEL AFFORDANCE of the THEME node. Since the value of the PATH attribute is functionally dependent on the value of the TRAVEL AFFORDANCE attribute, a restriction of the latter by the specification of one of its attribute values implies a restriction of the former. This dependency often leads to a value correlation between an attribute of the PATH node and an attribute of the TRAVEL AFFORDANCE node.

In addition to adverbs specifying velocity, the rule in (15) also allows for experiencer related modifiers such as *mühsam* ‘strenuously’ and *gemütlich* ‘comfortably’ as in the example repeated below.

- (16) *Der Weg steigt schnell / langsam / mühsam / gemütlich auf den Gipfel.*  
 the trail climbs quickly / slowly / strenuously / comfortably to the summit  
 ‘The trail climbs quickly / slowly / strenuously /comfortably to the summit.’

Modifiers of this type can be represented as values of the EXPERIENCE attribute of the TRAVEL AFFORDANCE node. As in the case of adverbs specifying values of VELOCITY, they are licensed because they can be interpreted as restricting the path. For instance, an adverb such as *mühsam* ‘strenuously’ can be conceived as being related to a steep SLOPE or a particularly meandering, non-linear SHAPE of the path. The way how the specification of the value of an attribute of travel affordance restricts the path also

---

<sup>14</sup>The “gentleness of the slope”/“a slow increase of elevation” as path properties being directly related to time adverbs such as *slowly* and likewise Japanese *yukkuri* ‘slowly’ has already been observed by Matsumoto (1996:202) with respect to fictive motion. We are grateful to one of the reviewers for pointing out to us that the alleged relation between velocity and slope does not necessarily have to hold (from a purely physical perspective). However, in our analysis we will keep with the prototypical relation between low velocity/gentle slope and high velocity/steep slope in accordance with observations such as the one made by Matsumoto.

seems to be influenced to some degree by the context. We leave it open here how the interaction between attribute-value pairs of the PATH and TRAVEL AFFORDANCE nodes can be captured in a formally adequate way.

As the attribute TRAVEL AFFORDANCE is naturally restricted to appear with entities which allow for travel, non-traversable referents do not come with this attribute. By consequence, modifiers such as *schnell* ‘quickly’ and *langsam* ‘slowly’, which specify a value of an attribute of TRAVEL AFFORDANCE, are excluded if *steigen*<sub>fict</sub> combines with a subject referent not suitable for human travel. As a result, the set of admissible modifiers found with non-travellable subject referents is considerably smaller in comparison to the array of modifiers attested in combination with travellable subject referents.

## 7 *Steigen*<sub>ins</sub>

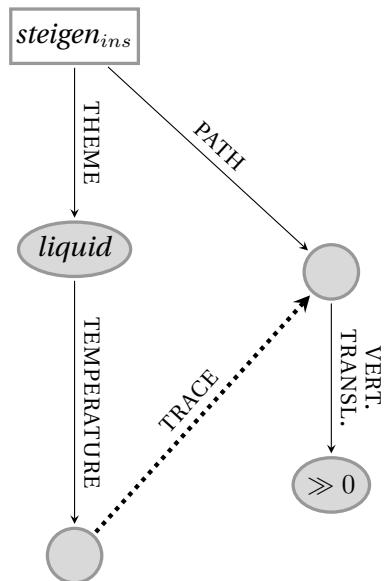
As illustrated by the example repeated in (17), the intensional reading is restricted to a positive value change, parallel to *steigen*<sub>fict</sub> and *steigen*<sub>dir</sub>.

- (17) *Die Temperatur steigt von 3 auf 10 Grad / \*von 10 auf 3 Grad.*  
 the temperature rises from to degrees from to degrees  
 ‘The temperature is rising from 3 to 10 degrees/\*from 10 to 3 degrees.’

Both, *steigen*<sub>ins</sub> and *steigen*<sub>fict</sub> are figurative readings. However, while the meaning of *steigen*<sub>fict</sub> remains in the same source domain ‘(geometrical) space’, *steigen*<sub>ins</sub> typically abstracts away into the domain denoted by the functional noun in subject position. Based on Gamerschlag et al. (2014) we assume the representation for *steigen*<sub>ins</sub> as in *Die Temperatur steigt* ‘The temperature is rising’ given in Fig. 10 below.

As can be seen, the frame of *steigen*<sub>fict</sub> is structurally nearly identical to the one of *steigen*<sub>dir</sub> except for the substitution of the POSITION-attribute by the TEMPERATURE-attribute. As with *steigen*<sub>fict</sub>, we consider this the result of an accommodation process triggered by a subject noun whose meaning is not compatible with one of the literal readings of the verb. However, as a contrast to *steigen*<sub>fict</sub>, this accommodation process embeds the meaning of the subject noun in a different way: Since the dimension that comes with the functional noun can be considered as an abstract value space, it is the POSITION-attribute which is targeted by this process, such that the geometrical value space is replaced by the particular abstract value space. Again, we assume that the value change which takes place during the *steigen*-event is recorded as a trace defined in terms of values with a temporal ordering. This trace is an abstract object which can be understood as a path through the value space determined by the particular dimension expressed by the functional noun in subject position, such as TEMPERATURE or PRICE. As with *steigen*<sub>dir</sub> and *steigen*<sub>fict</sub>, a positive value change is assured by restricting the values of VERTICAL TRANSLATION as being (considerably) greater than zero, with the difference that the values are determined to being e.g. TEMPERATURE-values or PRICE-values by the functional noun. Note that our paths

**Fig. 10** Frame representation of *Die temperatur steigt*. ‘The temperature is rising.’



are paths in an abstract value space. Thus the attribute VERTICAL TRANSLATION is not restricted to a spatial vertical difference but rather is a more abstract function which operates on intervals on the scale in focus (e.g., the temperature scale).

Note that the representation above does not refer to a stative scenario/fictive change, as a contrast to *steigen<sub>fict</sub>*. Rather, *steigen<sub>ins</sub>*, although abstracting away from geometrical space, is represented as an “ordinary” change in time resulting in a truly dynamic expression just like the one expressed by the near-synonymous change of state verb (*sich*) *erwärmen* ‘warm’.

## 8 Conclusion

In this paper, we have sketched how the fictive motion use of a verb such as German *steigen* ‘climb, rise’ can be systematically related to the dynamic readings of the verb by means of a frame analysis. Based on the observation that the intensional as well as the fictive motion use share with the directed motion reading the property that the value change expressed by the verb is restricted to a positive difference, we have argued that both figurative meanings are derived from the directed motion reading. Moreover, we have shown that both figurative uses trigger a different operation on the frame representation of the directional use: While the frame of the intensional use is derived from the one of the directional use by replacing the POSITION-attribute with the attribute that is specified by the subject noun, the fictive motion use is characterized by a deactivation of the dynamic components of the directed motion

meaning due to the stationary character of the subject referent. In the latter case, the meaning of the subject is accommodated as an entity embedding the (fictive) path of motion. The adverbial modifiers attested for this reading were shown to specify a property of the path related to a value of an attribute of the theme, either via value sharing or via covariation.

Since we have focused on a single verb of motion in one particular language in this paper, two strands of further research naturally arise. First, it is necessary to discuss more motion verbs, especially those which do not have a literal directional use as opposed to the manner use or vice versa. Additionally, a detailed corpus study would allow for the investigation of a broader array of modifiers which could serve as a probe into the precise meaning of the fictive reading. A particularly promising topic is the interplay between scalarity, telicity and dynamicity. Given that scalarity is independent from telicity and dynamicity (Fleischhauer and Gamerschlag 2014), the question emerges whether dynamicity and telicity are related. Usually, telicity is understood as a change until a specific endpoint/a specific degree on a scale is reached (e.g. Hay et al. 1999). If this is an adequate notion of telicity, telicity presupposes dynamicity. However, some change of state verbs, including German *steigen*, exhibit fictive motion uses which allow for modifiers indicating telicity such as the time-span adverbial *in kurzer Zeit* ‘within short time’ in (18) below.

- (18) *Die Straße steigt in kurzer Zeit um 200 Meter*  
 the road rises within short time by 200 meters  
 ‘The road rises by 200 meters within short time.’

The example above can be analyzed as spatially telic in the sense of Gawron (2009) and Champollion (2017) as an effect of adding the measure phrase *um 200 Meter* ‘by 200 meters’ whereas it can also be treated as ‘conventionally’ telic to some degree as indicated by the acceptability of the time-span adverbial *in kurzer Zeit* ‘within short time’. One central question to pursue in relation to these two different types of telicity is which role the simulative component of the representation plays in regard to the admissibility of the time-span adverbial and its telicity effect.

Second, the availability and flexibility of the fictive use of verbs of motion differs significantly crosslinguistically. For example, as already shown by Matsumoto (1996) for Japanese, the set of verbs available for the fictive motion reading can be confined in various ways. In particular, only verbs which highlight some aspect of the path of motion allow for a fictive reading, while verbs denoting the manner of motion are ruled out from this use. This restriction follows directly from Japanese being classified as a verb-framed language in which manner verbs cannot combine with spatial modifiers such as directional PPs and measure phrases. It needs to be clarified how this generalization can be implemented into the frame account above, which is not sensitive to this typological parameter. One technical way of addressing this aspect might be to exclude the value of PATH from the list of externally specified arguments for this class of verbs. However, we will leave it as an open question whether the satellite-versus verb-framed language distinction calls for a deeper representational asymmetry in both language types.

**Acknowledgements** The research presented in this paper was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft) with a grant to the Collaborative Research Centre (SFB) 991 “The Structure of Representations in Language, Cognition, and Science”. We are grateful to the two reviewers of this paper for many valuable comments. We would also like to thank the audiences of the CoSt16 conference and the Annual Event Semantics Meeting in Cologne for their feedback on an earlier version. Thanks also to Curt Anderson for commenting on and proofreading the final version of this paper.

## References

- Barsalou, L. (1992). Frames, concepts, and conceptual fields. In A. Lehrer & E. F. Kittay (Eds.), *Frames, fields, and contrasts* (pp. 21–74). Hillsdale NJ: Lawrence Erlbaum Associates Publishers.
- Champollion, L. (2017). *Parts of a whole: Distributivity as a bridge between aspect and measurement*. Oxford: Oxford University Press.
- Deo, A., Francez, I., & Koontz-Garboden, A. (2013). From change to value difference in degree achievements. In *Proceedings of SALT*, Vol. 23, pp. 97–115.
- Fillmore, Ch. (1982). Towards a descriptive framework for spatial deixis. In R. J. Jarvela & W. Klein (Eds.), *Speech, place and action: studies in deixis and related topics* (pp. 31–59). New York: John Wiley and Sons.
- Fleischhauer, J., & Gamerschlag, T. (2014). We’re going through changes: How change of state verbs and arguments combine in scale composition. *Lingua*, 141, 30–47.
- Gamerschlag, T., Geuder, W., & Petersen, W. (2014). Glück auf der Steiger kommt—A frame account of extensional and intensional ‘steigen’. In D. Gerland, Ch. Horn, A. Latrouite, & A. Ortmann (Eds.), *Meaning and grammar of nouns and verbs* (pp. 115–144). Düsseldorf: Düsseldorf University Press.
- Gawron, M. (2009). The lexical semantics of extent verbs. *Ms.*, San Diego State University.
- Geuder, W., & Weisgerber, M. (2008). Manner of movement and the conceptualization of force. *Talk presented at the workshop Il y a une manière et manière*. Université d’Artois, Arras, France. [http://semanticsarchive.net/Archive/Tk5YmEwN/MannerMovement\\_slidescompact.pdf](http://semanticsarchive.net/Archive/Tk5YmEwN/MannerMovement_slidescompact.pdf).
- Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, Acting and Knowing* (pp. 67–82). Hillsdale, USA: Lawrence Erlbaum.
- Gibson, J. J. (1979). The ecological approach to visual perception. New Jersey, USA, Lawrence Erlbaum Associates.
- Hay, J., Kennedy, Ch., & Levin, B. (1999). Scalar structure underlies telicity in ‘degree achievements’. In T. Matthews & D. Strolovitch (Eds.), *Semantics and Linguistic Theory (SALT) IX* (pp. 127–144).
- Hütte, St., & Matlock, T. (2016). Figurative language processing: Fictive motion and the visual world. In P. Knoeferle, P. Pyykkönen-Klauck, & M. W. Crocker (Eds.), *Visually situated language comprehension* (pp. 185–203). Amsterdam: Benjamins.
- Jackendoff, R. (1985). Multiple subcategorization and the theta-criterion: The case of ‘climb’. *Natural Language Linguistic Theory*, 3, 271–295.
- Kaptelinin, V. (2013). Affordances. The encyclopedia of human-computer interaction, 2nd Ed. <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/affordances>.
- Koontz-Garboden, A. (2010). The lexical semantics of derived statives. *Linguistics and Philosophy*, 33, 285–324.
- Kövecses, Z. (2015). *Where metaphors come from: Reconsidering context in metaphor*. Oxford: Oxford University Press.
- Langacker, R. W. (1986). Abstract motion. In *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*, pp. 445–471.

- Löbner, S. (1979). *Intensionale Verben und Funktionalbegriffe*. Tübingen: Narr.
- Löbner, S. (1981). Intensional verbs and functional concepts: More on the “rising temperature” problem. *Linguistic Inquiry*, 12(3), 471–477.
- Löbner, S. (2011). Concept types and determination. *Journal of Semantics*, 28(3), 279–333.
- Löbner, S. (2017). Frame theory with first-order comparators: Modeling the lexical meaning of punctual verbs of change with frames. In H. H. Hansen, S. E. Murray, M. Sadrzadeh & H. Zeevat (Eds.), *Selected papers of the 11th International Tbilisi Symposium on Logic, Language, and Computation*. LNCS, Vol. 10148, (pp. 98–117). Berlin: Springer.
- Montague, R. (1973). The proper treatment of quantification in ordinary English. In Hintikka, J., Moravcsik, J. & Suppes, P. (Eds.), Approaches to natural language. In *Proceedings of the 1970 Stanford Workshop on Grammar and Semantics* (pp. 221–242). Dordrecht: Reidel.
- Matlock T (2004a) The conceptual motivation of fictive motion. In G. Radden & R. Dirven (Eds.), *Motivation in grammar* (pp. 221–248). Amsterdam: John Benjamins.
- Matlock, T. (2004b). Fictive motion as cognitive simulation. *Memory and Cognition*, 32, 1389–1400.
- Matlock, T., & Bergmann, T. (2014). Fictive motion. In E. Dąbrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics* (pp. 771–790). Berlin: DeGruyter Mouton.
- Matsumoto, Y. (1996). Subjective motion and English and Japanese verbs. *Cognitive Linguistics*, 7(2), 183–226.
- Naumann, R. (2013). Outline of a dynamic theory of frames. In G. Bezhaniashvili, S. Löbner, V. Marra & F. Richter (Eds.), *Logic, language, and computation*. LNCS, Vol. 7758 (pp. 115–137). Berlin, Heidelberg: Springer.
- Petersen, W. (2007). Decomposing concepts with frames. *Baltic International Yearbook of Cognition, Logic and Communication* 2 (pp. 151–170). Reprint 2015. In T. Gamerschlag, D. Gerland, R. Osswald & W. Petersen (Eds.), *Meaning, frames, and conceptual representation* (pp. 43–67). Düsseldorf: Düsseldorf University Press.
- Talmy, L. (2000). Toward a Cognitive Semantics, Vol. I. Concept structuring systems. Cambridge (Mass.): MIT Press.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Cascades. Goldman's Level-Generation, Multilevel Categorization of Action, and Multilevel Verb Semantics



Sebastian Löbner

**Abstract** The paper proposes a novel theory of the categorization of acts and applies it to the semantics of action verbs, with fundamental consequences for semantic theory and beyond. The theory is based on Goldman's (*Theory of human action*. Princeton University Press, Princeton, NJ, 1970) multilevel theory of action which is taken here as a theory of categorization. Goldman's central notion is *level-generation*: acts of a type may under circumstances generate acts of other, more abstract types. The acts form a hierarchical structure which Goldman calls an *act-tree*. Level-generation results in a conceptual relation called *c-constitution* here, i.e. constitution under the given circumstances; I also introduce the more general term *cascade* for act-trees. In the second part, multilevel cascade-structure categorization is combined with a cognitive semantics that models meanings with Barsalou frames. A multilevel analysis of the concept of writing is discussed in depth and detail in order to illustrate the potential and the consequences of a cascade approach to verb semantics. It is shown that the concept of c-constitution can be generalized as to cover the roles of persons and objects across levels in a cascade. The generalization suggests that multilevel categorization may be a very general and fundamental phenomenon in the psychology of categorization.

**Keywords** Level-generation · C-constitution · Cascades · Multilevel categorization · Frames · Decomposition · Action verbs · Composition · Reference · Ontology

---

S. Löbner (✉)

Institute for Language and Information, Heinrich-Heine-Universität, 40204 Düsseldorf, Germany  
e-mail: [loebner@phil.hhu.de](mailto:loebner@phil.hhu.de)

## 1 Introduction

### 1.1 *The Intuitive Notion of “Level-Generation”*

Our point of departure is a philosophical theory from as far back as 1970, the year when the first seminal papers by Richard Montague appeared and triggered the development of formal semantics. Goldman’s theory of “level-generation” was the first general theory of action<sup>1</sup> to come up with the idea (and observation) that we consider ordinary tokens of acts very often as representing more than one type of act. While it is an almost trivial fact about categorization that one and the same thing can always be categorized in numerous different ways, Goldman’s theory makes a much stronger claim: His basic mechanism of “level-generation” relates multiple categorizations of the same doing in *systematic* ways. Under given circumstances, level-generation yields a whole tree of categorizations, such that doing a particular thing amounts to, or constitutes, doing at the same time—in one—a variety of things of different types. Goldman emphasizes that his notion of level-generation meets a basic intuition, and you will see that it does from just a handful of examples (in (1) in the box). These examples are to be read as follows: start from the bottom and follow the ↑ arrows; these symbolize level-generation. Assume that for each example the given circumstances are such that they allow to read the arrow as “and thereby”, or “this constitutes”. You can easily imagine (or reconstruct) circumstances that would support these steps of level-generation. The vertical structures are trees; for the sake of simplicity, the trees in (1) don’t branch, but you will see below that trees can. The trees consist of acts by the same agent and they coalesce acts that are all done in one: x, in one, flips the light switch, turns on the light, lightens the room, wakes the baby, ruins their night—all done by one little movement of a finger. The same holds for all other examples of level-generation. Being done in one, all those acts in a tree happen at the same time.

---

<sup>1</sup>In fact, Austin’s speech act theory anticipated Goldman’s multi-level approach, but it was not applied beyond the special subclass of acts constituted by speech acts. We will give due credit to Austin’s speech act theory in Sect. 6.1.

## (1) Examples of level-generation

a.

$x$  ruins their night  
 $\uparrow$   
 $x$  wakes the baby  
 $\uparrow$   
 $x$  lightens the room  
 $\uparrow$   
 $x$  turns on the light  
 $\uparrow$   
 $x$  flips the light switch

b.

$x$  makes  $y$  smile  
 $\uparrow$   
 $x$  does  $y$  a favor  
 $\uparrow$   
 $x$  lets  $y$  pass  
 $\uparrow$   
 $x$  keeps the door open

c.

$x$  purchases  $z$   
 $\uparrow$   
 $x$  pays for  $z$   
 $\uparrow$   
 $x$  hands money to  $y$

d.

$x$  disappoints  $y$   
 $\uparrow$   
 $x$  declines  $y$ 's request  
 $\uparrow$   
 $x$  says "No" to  $y$

e.

$x$  runs a personal record over 100 m  
 $\uparrow$   
 $x$  runs 100 m in 12.3 seconds

These examples all seem natural. Without much reflection, we would agree that in all these cases the upward arrow may, under appropriate circumstances, be expressed as “and thereby” and always means the same; and it is natural to view these examples as different types of act done in one. It is this intuitive connection between different ways in which—under circumstances—a given act can be categorized that Goldman’s theory of action is about.

Level-generation is an extremely common thing. If we think of it, we realize that our minds are doing it automatically and inevitably all the time. If somebody does something concrete, we will categorize it not just as a basic bodily action like keeping a door open, handing money to somebody, or pressing a button. We will rather have our attention on what the person is doing *thereby*, because what will matter to us will not be the mere bodily movements, meaningless in themselves, but what they achieve (or try to achieve). The same applies to our own actions and the ways we *mean* them. We don’t mean to exercise our thumb, when we press a button on the remote control—we mean to turn on the TV. Most, if not all, things we physically do we do not do just for themselves.

## 1.2 *The Structure of the Chapter*

Goldman originally presented his theory as a contribution to philosophical ontology. He argued that under circumstances like those assumed in the examples, the agent exemplifies multiple different acts in one. Not every ontologist would follow him; many would argue that the agent does just one thing which may happen to meet different descriptions, under circumstances.

I will re-construe Goldman's theory not as an ontological theory of action, but as a theory of the cognitive *categorization* of action, a view which Goldman actually supported later in arguing that the notion of level-generation is “a psychological structure, or the manifestation of a psychological structure” (see (7) in Sect. 2.3 for the full quote from Goldman 1979). This turn has important consequences. First, Goldman's theory is turned into a theory of cognitive representation, and his mechanism of level-generation receives the role of a *cognitive* mechanism. Second, it makes the theory immune to the ontological objection that there exist only one doing, not several distinct ones: the fact that one doing may, under circumstances, be categorized in multiple ways, is uncontroversial. Third, the psychological turn makes Goldman's theory applicable to linguistic semantics (of a cognitive orientation); as you will see, it is to be assumed that level-generation is written into the lexical meanings of probably almost all verbs of action.

In Sect. 2, I will briefly review Goldman's original theory and its reception in the philosophical discussion. My own construal of the theory will be made precise; I will introduce the central notions of ‘cascade’ and ‘c-constitution’ replacing Goldman's ‘act-tree’ and ‘level-generation’, respectively. Section 3 provides examples and data that illustrate the relevance of level-generation for verb semantics and verb grammar.

The second part will be concerned with a formalization of c-constitution and cascades in the framework of Düsseldorf Frame Theory and the application of the approach to semantics. In Sect. 4, act-cascades will be modeled as trees of first-order frames that each represent a single type of action (like ‘flip the light switch’ or ‘wake the baby’). Section 5 will treat in depth an illustrative, more complex example, the ‘write’ cascade. I will discuss the far-reaching consequences of a cascade approach to action verb meanings for theories of lexical meaning, composition, and reference in Sect. 6. The chapter will be concluded with a brief reflection of the perspectives that the multilevel approach to categorization opens up for cognition, semantics, and life.

## 2 Level-Generation: Doing Multiple Things in One

### 2.1 Preliminary: *Act-Tokens, Act-Types, and Act-TTs*

The upward relation symbolized by the arrow ↑ in the examples represents what Goldman called level-generation. The first question concerning this notion is: what kind of thing does it relate. Goldman (1970) distinguishes act-tokens and act-types. *Act-types* are common enough: it is types such as ‘open the door’, ‘turn on the light’, ‘wake the baby’, or ‘decline a request’.<sup>2</sup> They can be defined more or less specifically, for example as ‘open’, ‘open a door’, ‘open (a particular) door’, ‘x open (a particular) door’ etc. In philosophy, types of act (or action) are often subsumed

---

<sup>2</sup>Descriptions of types will be marked by single quotes.

under the notion of “property”, in semantics, under “types of events”. Act-types are exemplified/enacted/performed/implemented if someone does something of that type. The agent then produces an *act-token* of this type. If Sue does something that can be described as “open the door”, she produces a token of the act-type ‘open the door’. An act-token has a determinate agent and occurs at a determinate time.

According to Goldman’s approach, level-generation obtains between act-tokens in this sense; there is a token of ‘flip the light switch’ that level-generates a token, by the same agent and at the same time, of ‘turn on the light’, and so on. In Goldman’s account, two act-tokens are different if they are tokens of different types, and two tokens are only identical, if they are tokens of the same type; more precisely:

- (2) a. “Each act-token is a token of one and only one type (property).” (Goldman 1970: 11)
- b. “Two act-tokens are identical if and only if they involve the same agent, the same property, and the same time.” (Goldman 1970: 10)

Thus, according to him, the tokens in one act-tree are distinct. The conditions in (2) mean that the relation of level-generation does not obtain between act-tokens as such, but between acts-as-tokens-of-a-type. For example, (1d) is to be construed as: a token of the act-type ‘say “No” to y’ level-generates a token of the act-type ‘decline y’s request’, and this in turn a token of the act-type ‘disappoint y’.

Tokens-of-a-type are a very natural kind of thing. Whenever we talk about acts or events, we do so while describing them as of one type or another. For example, if we use a VP for event reference, the VP provides a description of the event referred to and thereby gives its type. Language cannot refer to acts other than by type description and semantic and pragmatic means that fix the reference to particular tokens of *that type*. This does not only hold for acts and events, but in general for all things we verbally refer to: we always refer qua type, that is, using expressions that provide a type description. It may even be argued that this applies beyond language to thinking in general: we can’t think of things, or even perceive things, without categorizing them in one way or another.

I will refer to a token-of-a-type as a “TT” for short, and introduce the following notation:

- (3) **Definition:** For a type T and an entity t,  $t/T$  is the “token t of the type T”.

TTs are essentially ordered pairs of an entity and a type such that the entity is of this type. It follows immediately that two TTs  $t/T$  and  $t'/T'$  are different if T and T’ are. Goldman himself never speaks explicitly of act-tokens-of-a-type, but always of act-tokens and of act-types. However, due to the conditions in (2), he implicitly talks of TTs whenever he talks of act-tokens in the context of his theory. We will keep this in mind for the following discussion.

## 2.2 Goldman's Theory of Act-Levels

### 2.2.1 The Multilayered View on Human Action

Goldman's point of departure is the observation that agents when they act may do several distinct things in one; they produce a set of several act-tokens. Goldman emphasizes that these act-tokens are distinct "because", he argues, "the *properties* picked out [...] are distinct properties" (Goldman 1970: 12, his italics)—flipping the light switch is not a token of the same property as turning on the light is a token of, etc. One crucial difference of the properties distinguished concerns the respective causal relationships of the types of action: flipping the light switch may cause the light to go on, but turning the light on does not cause the light switch to be flipped. As a consequence of the regulations in (2), acts related to each other like in the examples cannot be identical as they are tokens of different properties. Goldman presents this argument against the proponents of what he calls the "identity thesis" put forward by Anscombe (1963) and Davidson (1963), among others he mentions [p. 2]. According to Goldman, there is one doing by the agent that constitutes a combination of distinct act-tokens of distinct act-types. Our construal of Goldman's—that he is actually talking of TTs—avoids the ontological controversy between "unifiers" (Davidson, Anscombe and others) and "multipliers" (Goldman himself).<sup>3</sup>

### 2.2.2 Act Levels and Level-Generation

In Goldman's theory of action, the act-tokens enacted with a single doing are ordered in levels. Act-tokens at lower levels "level-generate" higher-level act-tokens of the same agent at the same time. If an act-token  $a$  by agent  $s$  level-generates an act-token  $a'$ , then  $s$  does  $a'$  "by" or sometimes "in" doing  $a$  [pp. 20–1]. Goldman distinguishes four general types of level-generation. One of them is "augmentation generation"; I will set it apart from the other three (as Goldman himself does, to a degree) and turn to it later in Sect. 2.5. I will use original examples from Goldman (1970) in order to introduce and illustrate Goldman's types of level-generation. As above, I use the symbol  $\uparrow$  for level-generation, but I do not yet apply the notion of act-TTs, as I want to quote Goldman's original definitions. A restatement of Goldman's notions in terms of TTs will be undertaken in Sects. 2.5 and 2.6.<sup>4</sup>

---

<sup>3</sup>Ginet (1990) devotes a chapter to the question whether or not the acts in an act-tree are identical or not and comes to the conclusion that "the issue over the individuation of action, though sufficiently interesting in its own right, is not one on which much else depends. As far as I can see, there is no other significant question in the philosophy of action that depends on it." [p. 70].

<sup>4</sup>In the quotes, I replace the original upper-case letters for variables denoting act-tokens and persons by lower-case letters, as I want to reserve in this paper the use of upper-case letters for type variables.

(4) 1. **Causal generation**

"Act-token  $a$  of agent  $s$  causally generates act-token  $a'$  of agent  $s$  only if  
 (a)  $a$  causes [an event]  $e$ , and  
 (b)  $a'$  consists in  $s$ 's causing  $e$ ." [p. 23]

Goldman's examples [p.23]:

- ' $s$  flips the switch'  $\uparrow$  ' $s$  turns on the light'
- ' $s$  shoots the gun'  $\uparrow$  ' $s$  kills George'
- ' $s$  closes the door'  $\uparrow$  ' $s$  prevents a fly from entering the house'.

Among the introductory examples, (1a) and (1b) involve causal generation in all steps. In order to avoid confusion, it is very important to keep in mind that causal level-generation does not relate an act  $a$  with an event  $e$  caused by  $a$ , but an act  $a$  with the act  $a'$  of causing such an event. For example, it does not relate the act of turning on the light with the event of the baby waking up; rather it relates the act of turning on the light with the act of waking the baby. Unlike the other two types to follow, causal generation raises the question as to whether the generating and the generated act happen at the same time. Goldman points out [p. 21] that it is generally inadequate for two acts  $a$  and  $a'$  related by level-generation to state that the agent did  $a$  and then did  $a'$ . This holds even if  $a'$  is causally generated and the effect caused sets in only later than  $a$  is done; thus, even if in the case of, say, (1d)  $y$  learns of  $x$ 's declining  $y$ 's request only several days later, one would not say that  $x$  declined  $y$ 's request and then disappointed her. Rather the disappointing act was done when  $x$  declined the request.

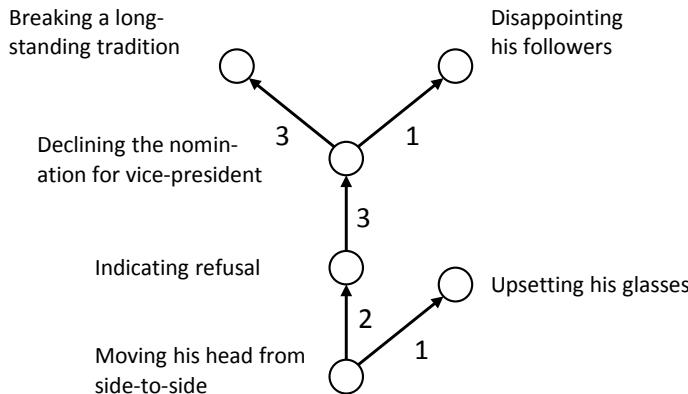
[4)] 2. **Conventional generation**

"Act-token  $a$  of agent  $s$  conventionally generates act-token  $a'$  of agent  $s$  only if the performance of  $a$  in circumstances  $c$  (possibly null), together with a rule  $r$  saying that  $a$  done in  $c$  counts as  $a'$ , guarantees the performance of  $a'$ ." [p. 26]

Goldman's examples [p. 25]:

- ' $s$  moves his queen to king-knight-seven'  $\uparrow$  ' $s$  checkmates his opponent'
- ' $s$  breaks his promise'  $\uparrow$  ' $s$  does what he ought not to do'
- ' $s$  extends his arm out the car window'  $\uparrow$  ' $s$  signals for a turn'

(1c) is a case of conventional generation; in (1d), the first step is conventional, the second is causal.



**Fig. 1** Goldman's act-tree for declining the nomination for vice-president

### [(4)] 3. Simple generation

"In simple generation the existence of certain circumstances, conjoined with the performance of a, ensures that the agent has performed a'." [p. 26]

Examples [p. 27]

- 's jumps 6 feet 3 inches'   ↑   's outjumps George'
- 's comes home after 12:00'   ↑   's breaks his promise'
- 's asserts that p'                 ↑   's lies'

The distinction of types of level-generation reflects the fact that level-generation may draw on different types of connection between actions: on causal connections, on convention, or just on the constellation of facts (simple generation).

Goldman uses "act-tree" diagrams for complex level-generational act structures; the trees are to be read bottom-up. The act-tree in Fig. 1 contains instances of all three types of level-generation listed in (4).<sup>5</sup> The diagram displays six nodes that stand for act-tokens of different types as labeled. They are connected by arrows indicating the direction of generation. The numbers indicate the three types of level-generation as numbered in (4). The tree contains two act-nodes with upward branching generation. Moving the agent's head not only conventionally generates indicating a refusal, but also causally generates upsetting the agent's glasses. The agent's declining the nomination causally generates his disappointing his followers; it also generates in simple generation breaking a long-standing tradition. The latter constitutes simple generation because it comes about by the mere circumstances of such a tradition having obtained for a long time. If an act-token generates two or more others which do not generate each other, the generated acts are both at a higher level, but the levels are independent of each other; in particular, they are not the same level. According

<sup>5</sup>The diagram is adapted from Goldman (1970: 34), with dots replaced by circles, and lines by upwards arrows. I omit the first step of the act-tree as it consists in augmentation generation.

to Goldman [p. 31], two acts are “at the same level” if and only if they are distinct but generated by the same act and generating the same acts. His examples include ‘hitting the tallest man in the room’ and ‘hitting the wealthiest man in the room’ where in the circumstances given the tallest man in the room happens to be the wealthiest one. I will neglect the issue of same-level acts in the following.

Goldman gives the following general definition of level-generation.<sup>6</sup> He also includes the type of augmentation generation which we exclude, but the definition applies to the three types in (4) just the same.

- (5) “Act-token a level-generates act-token a’ if and only if
  - (i) a and a’ are distinct act-tokens of the same agent that are not on the same level;
  - (ii) neither a nor a’ is subsequent to the other; neither a nor a’ is a temporal part of the other; and a and a’ are not co-temporal;
  - (iii) there is a set of conditions c\* such that
    - (a) the conjunction of a and c\* entails a’, but neither a nor c\* alone entails a’;
    - (b) if the agent had not done a, then he would not have done a’;
    - (c) if c\* had not obtained, then even though s did a, he would not have done a’.”

The condition in (ii) that a and a’ be not co-temporal is in need of explanation. According to Goldman’s introduction of the term, two acts a and a’ are “co-temporal” if and only if the agent of a does a “while also” doing a’, as an instance, one might add, of multitasking. If x turns on the light by flipping the light switch, x does not flip the light switch while also turning on the light. Thus, condition (ii) bars level-generation between acts exerted in parallel. It does not preclude that the acts related by level-generation do not have the same temporal extension—to the contrary, they necessarily have. “There is a sense [...] in which pairs of generational acts are always done *at the same time*” Goldman explains [pp. 21–2].

Goldman’s definition captures important basic properties of level-generation<sup>7</sup>:

---

<sup>6</sup>P. 43, italics omitted, Arabic numbering replaced by Roman, upper-case variables by lower case.

<sup>7</sup>Another general characterization of level-generation is to state that it is a supervenience relation: the generated act supervenes the generating act. McLaughlin and Bennett (2014) give the following definition: A set of properties A supervenes upon another set B just in case no two things can differ with respect to A-properties without also differing with respect to their B-properties. Supervenience is a very weak correspondence relation, while level-generation is much more specific. To state that level-generation is a supervenience relation does not mean to say that it is *merely* supervenience.

- (6) Basic properties of level-generation
- a. Generating act and generated act are acts by the **same agent**.
  - b. Generating act and generated act have the **same temporal extension**.
  - c. Level-generation is a **dependence relation**:  
Generated acts depend on the generating act, and appropriate circumstances, to come about.
  - d. The types of the generating act and the generated act are **logically independent**:  
In principle, when an act of the generating type is exerted, there need not be an act of the type generated, and vice versa.

Goldman's definition secures the basic relational properties of level-generation. The relation of "level-generation is intended to be asymmetric, irreflexive, and transitive" (Goldman 1970: 22). Since it is irreflexive, no act generates itself. Asymmetry prevents two acts from generating each other. Due to transitivity, if a generates b and b generates c, then a generates c. As a consequence of transitivity, level-generation may result in chains, and due to irreflexivity and asymmetry the chains cannot contain loops. (If loops are not excluded, acts in a loop would generate themselves and generate their generators.)

Transitivity has two important consequences. First, we may combine a given sequence of level-generations into one larger step. For example in (1a) we might skip some of the levels; somebody might warn the agent: "if you flip this switch, you'll ruin your night!" Second, it may conversely be possible that a given step be broken down into several smaller steps. For instance, one might analyze the level-generation of 'flip the light switch' ↑ 'turn on the light' into more steps that take into account what the agent does on the mechanical and the electrical level, like closing an electric circuit and thereby providing electricity to the bulb in a lamp, heating a wire and making it radiate light. A fine-grained analysis like this might matter under circumstances where the attempt to turn on the light by flipping the switch fails.

Asymmetry, irreflexivity, and transitivity hold for generalized level-generation comprising the causative, conventional, and simple type. It is these logical properties of level-generation that give rise to tree structures as the one in Fig. 1.

## 2.3 *Critics of Goldman's Theory*

Goldman's theory was criticized by Castañeda (1979), Bennett (1988), and McCann (1982), among other philosophers. The central target of criticism is Goldman's formal definition of level-generation quoted in (5). The critics showed by counterexamples that it would apply to cases of act pairs that are obviously not intended to be included. This criticism is justified, but it fails to invalidate Goldman's theory of level-generation; it just shows that Goldman's attempt at a formal definition did not achieve an adequate description of level-generation.

Goldman's definition in (5) is essentially in terms of logical conditions on two statements *s does a* and *s does a'* where *s*'s doing a level-generates *s*'s doing *a'*. Logical conditions, properties, and relations are in terms of truth-values (entailment) or in terms of extensions of concepts. For example, if a sentence *B* is always and necessarily true if sentence *A* is, then *A* and *B* are related by logical entailment: *A* entails *B*. If a concept *P* is such that it applies to all cases that another concept *Q* applies to, then *P* is in the logical relationship of superordination to *Q*. By contrast, conceptual relations concern the conceptual content. For example, the two sentences *Today is Tuesday* and *Tomorrow is Wednesday* logically entail each other, but they are not the same. There are conceptual meaning relations between them that *explain* why they are logically equivalent (both refer to a day, the second sentence to a day following the one referred to in the first; Wednesdays are related to Tuesdays in the same way). Logical relations derive from conceptual relations; for example it derives from the concepts of 'perceive' and 'hear' that 'x hears y' logically entails 'x perceives y'. But conversely, no particular conceptual relation derives from entailment. Thus, Goldman's condition (5iiia) does not tell us how the categorizations of *a* and *a'* are conceptually related, for example in the way that *a'* of type *A'* is done by exemplifying some *a* of type *A*. Taking a look at the conditions in (5), we realize that (5i) is just a restricting precondition for the definition, and that the conditions in (5iii) are in terms of logical entailment (or can be paraphrased as such). The only (probably) non-logical condition is the restriction in clause (5ii) that *a* and *a'* be not co-temporal; but this weak constraint is far from capturing the basically non-logical notion of level-generation. Level-generation, as introduced by Goldman, is a genuinely conceptual, or as I see it, cognitive relation. In his reply to Castañeda (1979), Goldman explicitly locates level-generation in the realm of psychology:

- (7) “[...] insofar as philosophical theorizing is an attempt to lay bare the fundamental features of our conceptual scheme [i.e. level-generation, S.L.], it should not rest content with a “string” of explicit definitions. Our conceptual scheme is a *psychological structure, or a manifestation of a psychological structure*, and it is not the analysis of concepts alone that will facilitate our understanding of this structure.” [Goldman 1979: 269, my italics]

Given that, Goldman's definition in (5) fails to capture the real nature of the notion of level-generation—in fact no definition in terms of logical relations can. A definition like the one intended in (5) can only provide necessary logical conditions to be met by level-generation. The critics mentioned were right in pointing out that Goldman's attempt at a [logical] analysis of the relation does not provide a sufficient condition; but this circumstance does not invalidate the underlying intuitive notion of level-generation that Goldman's attempt at an analysis was aimed at.

- (8) “[...T]he idea of level-generation, I think, is an intuitive or pre-analytic idea, implicit within our common-sense framework. [...T]he idea of level-generation is implicit in our use of the phrase, “s did ... *by* doing —,” and in our use of the phrase, “s did ... *in* doing —.” That it is an intuitive notion is reflected in the fact that once a few examples of it are given, any ordinary speaker can readily identify numerous other cases that fall under the same concept. [...] Since there is a prior notion to be analyzed, we do not want to provide merely a *stipulative* definition. We want to provide a definition that captures our antecedent notion (while also capturing the amplifications of the notion – e.g., augmentation generation – which I have introduced). But providing analyses of interesting concepts is always a difficult enterprise. What must be remembered, therefore, is that the tenability of the intuitive concept should not depend on the success of any particular analysis.” [Goldman 1970: 38]

It appears uncontroversial to consider the rich analysis of doings like the ones indicated in the examples as “real” in the sense that if an agent acts in a particular situation and we consider a multilevel conceptualization adequate, then all the act-types, to us, are “really” enacted in this one doing. Thus, Goldman’s theory of human action can be considered a contribution to ontology, and metaphysics, of the world *as it is perceived and conceived by human cognitive agents*, i.e. of what is *real to us*.

## **2.4 Goldman’s Theory of Human Action Applied to Cognitive Representation**

In view of the two quotes cited, I will apply Goldman’s theory to the cognitive *representation* of human action (a construal which was not applied by the philosophical critics). If, to us, an act constitutes a whole tree of act-TTs, I will assume that our cognitive representation has this tree structure, composed of representations of the participating types of act. I assume that level-generation is a fundamental cognitive mechanism, ubiquitously at work in our cognitive systems. Whenever somebody acts, we will try to interpret their action at levels beyond the pure doing, and will thereby come up with a view that, for example, explains the action as the result of the agent pursuing certain intentions to be accomplished at some level generated; we will try to relate the action to ourselves as some type of act towards us; we will often appraise the action as positive or negative in various regards; we will take it as constituting *interaction* with ourselves, and so on. All these views amount to the addition of cascade levels to the doing. Thus, there are quite general level-generations we may assume, like the following:

- (9) a.  $x \text{ does } a/A \uparrow x \text{ does } b/$  ‘pursue intention Y’ intentionality  
 b.  $x \text{ does } a/A \uparrow x \text{ does } b/$  ‘do sth. I like or dislike’ appraisal  
 c.  $x \text{ does } a/A \uparrow x \text{ does } b/$  ‘direct a/A at me’ interaction  
 d.  $x \text{ does } a/A \uparrow x \text{ does } b/$  ‘prepare a situation of type S’ sequentiality  
 e.  $x \text{ does } a/A \uparrow x \text{ does } b/$  ‘react to a situation of type S’ sequentiality

In view of such examples, it is hard to imagine that we do *not* level-generate whenever we observe the actions of others, or plan and execute our own. Level-generation as a cognitive process will very often be automatic, not involving any conscious reasoning.

Construing Goldman’s as a theory of cognitive representation of action will enable us below to apply it to semantics—which I take to be part of a theory of cognitive representations, too, in this case of linguistic meanings. But before we turn to this aspect, I will restate the basic points of the theory in terms of act-TTs, and also undertake a slight revision of Goldman’s view of “augmentation generation”.

## 2.5 *Level-Generation and Augmentation Generation*

Goldman (1970: 28–30) distinguishes three subtypes of what he calls “augmentation generation”<sup>8,9</sup>:

(10) **Subtypes of augmentation generation**

a. **Compound augmentation** [our term]

Two or more acts by the same agent and at the same time (“co-temporal” acts) jointly generate an act of doing all these things in one.

Ex. ‘s jumps’, ‘s shoots’ generates ‘s jump-shoots’ [p. 28]

b. **Manner augmentation** [our term]

An act generates doing this act in a particular manner.

Exx.: ‘s says “hello”’ generates ‘s says “hello” loudly’  
 ‘s runs’ generates ‘s runs at 8 m.p.h.’ [p. 28, 29]

c. **Argument augmentation** [our term]

An act generates another act distinguished by the specification of an additional argument.

Exx.: ‘s extends his arm’ generates ‘s extends his arm out the car window’  
 ‘s moves his queen’ generates ‘s moves his queen to king-knight-seven’ [p. 34]

---

<sup>8</sup>For the sake of terminological consistency, I replace Goldman’s original term ‘compound generation’ by ‘compound augmentation’.

<sup>9</sup>The term ‘argument’ is used here in a sense also including adjuncts.

Goldman himself did not seem entirely convinced that augmentation generation is of the same kind as the other three types of level-generation (cf. his discussion pp. 28–30). Related to the conceptual level, augmentation in all varieties mentioned is enrichment of a given act-type concept: the original concept is maintained and a condition, or circumstance, added such as to form a concept that is more specific. In ‘extend one’s arm *out the car window*’, the direction of the movement is added as a particular circumstance, analogously for manner augmentation; for compound augmentation, the co-temporal acts constitute the crucial circumstances for each other.

The application of the augmented concept must be narrower than the application of the concept augmented. If a concept  $A+$  is an augmentation of a concept  $A$ , then  $A+$  unilaterally entails  $A$ , that is,  $A$  applies to all cases to which  $A+$  applies, but not conversely. As we saw in (6d), entailment does not pertain with the other types of level-generation.

Rather than attempting to subsume augmentation under level-generation, I recognize the conceptual process as a mechanism of its own, independent of the phenomenon of level-generation. Augmentation is the well-known, basic, and ubiquitous conceptual process of concept enrichment: a given concept/categorization/type is enriched by adding conditions. Thereby the extension of the concept is narrowed down. As a cognitive process, augmentation, or enrichment, is of fundamental importance. It underlies learning in form of gradual differentiation of a concept; it is involved in all processes of adding information to existing knowledge representations, including concepts for categories. In the theory of types such as in Carpenter (1992), the relationship between a given type and an enrichment of it is established as “subsumption”, the wider, less rich, type subsumes the narrower, enriched type.

Augmentation is a basic process *along with* level-generation; it may even be more general. The definition in (11a) defines the general notion as a relation between concepts in general; it applies to act-types in particular. The definition is generalized in (11b) as to cover Goldman’s compound augmentation. (11c) defines the derived notion of an act-TT  $a+/A+$  being more specific than an act-TT  $a/A$ ; in the case of compound augmentation, the relation holds between each component act and the compound act.

### (11) Augmentation

- a. A concept  $A+$  is an augmentation of the concept  $A$ , or:  
 $A$  properly subsumes  $A+$   
 $A \sqsubset A+$   
 iff  $A+$  is  $A$  with conditions added such that there are cases where  $A$  applies, but not  $A+$ , while  $A$  always applies if  $A+$  applies.
- b. For  $n > 1$ , the concept  $A+$  is an augmentation of the concepts  $A_1, \dots, A_n$ ,  
 $A_1, \dots, A_n \sqsubset A+$   
 iff  $A+$  is an augmentation of each act concept  $A_1, \dots, A_n$ .
- c. An act-TT  $a+/A+$  is **more specific** than an act-TT  $a/A$ , iff  $A \sqsubset A+$ .

By referring to the act tokens as “a” and “a+”, it is not implied that they are different as such. In fact, by the very definition, if a+ is a token of act-type A+, then it also is a token of all act-types A that subsume A+. The notation for the act tokens is chosen for convenience in order to fit in with the distinction of act tokens involved in c-constitution. We will refer to both, the relation between types and the relation between TTs, as augmentation.

Augmentation shares certain basic properties with level-generation. (i) By definition, augmentation preserves all information. Thus, if we apply augmentation to an act-TT a/A, then the agent of a+/A+ is necessarily the same as the agent of a/A; the same holds for the act times of a and a+. Note that this also holds in the case of compound augmentation: the subsumption relation can only obtain between A<sub>1</sub>, ..., A<sub>n</sub> and A+ if all n + 1 act-types have the same agent and time specification. Thus, the analogue of (6a, b) applies to augmentation. (ii) Augmentation, too, is an asymmetric, irreflexive, and transitive relation between act-TTs, and hence generates tree structures. Applied in the same domain, we can form trees that involve both augmentation and level-generation. However, there is one fundamental difference between augmentation and level-generation in the narrower sense: level-generation requires logical independence, while augmentation involves logical entailment.

I define “cascades” basically as Goldmanian act trees. I introduce a new term because I want to be able to extend the notion to multilevel representations of things other than acts.

### (12) **Act cascades**

An act cascade is a tree structure of act-TTs that are related by (causal, conventional, or simple) level-generation and/or by augmentation.

According to this definition, act-cascades are co-extensive with Goldmanian act-trees, but they are considered to be not all produced by sub-types of what *I* call “level-generation”.

## **2.6 C-Constitution**

### **2.6.1 The Relations c-by and c-in**

Goldman mentions the two options of paraphrasing the downward relationship between a generated act-TT h/H and its generator l/L, with a *by* or an *in* paraphrase: ‘Agent does h/H *by* doing l/L’ or ‘Agent does h/H *in* doing l/L.’<sup>10</sup> He exempts augmentation. Goldman does not elaborate on the question as to when one or the other type of paraphrase is adequate, but there is some discussion in Kearns (2003), although she does not refer to Goldman’s theory. Kearns discusses *in* versus *by* paraphrases in connection with certain action predicate types, to be discussed in Sect. 3.3

---

<sup>10</sup>I will use ‘L’, ‘L1’, ‘L2’, ... for lower cascade levels, and ‘H’, ‘H1’, ‘H2’, ... for higher levels.

as “criterion predicates”. What I refer to as lower and higher level, she calls ‘host’ and ‘parasite’, respectively. According to her, an *in* paraphrase expresses that “the host simply realizes the parasite” [p. 602]; while a *by* paraphrase expresses that “the causative parasite is not realized simply in the occurrence of the one action performed, but requires also a consequential upshot” [p. 615]. It is not clear from her discussion either, when which of the two paraphrases applies. Still, Kearns’ observation that the *in* paraphrase applies when the generating act *simply realizes* the generated act seems to be a valid generalization. We would say, for example, in the case of (13) that the casting of the speaker *is* the mistake.

- (13) *All through The Graduate Nichols thought he'd made a mistake in casting me.*  
 [BNC C9U 495]

By contrast, cases of generation where a *by* paraphrase is adequate seem to not allow for the equation, in this sense, of generating and generated act:

- (14) *Our aim is to reduce the number of new HIV infections by giving young people the facts about AIDS and by encouraging them to think about their future.* [BNC A01 532]

Clearly, giving young people the facts about AIDS *is* not, in itself, a reduction of the number of HIV infections, rather it is a possible *means*, or *method*, of achieving that. I conclude that there are two distinct inverse cascade relations that can be described by using *in* or *by*, respectively. These are alternative inverses of the relation of level-generation. I index the relations with the subscript ‘c’ for the given circumstances since these relations, like level-generation, only hold under circumstances.

- (15) The downward relation **c-in**

h/H **c-in** l/L, iff

Under the given circumstances c,

- the agent, **in** doing l/L, exemplifies an act h of type H;
- doing h/H **consists in** exemplifying an act l of type L;
- the agent’s doing l/L **counts as / amounts to / means** exemplifying an act h of type H.

- (16) The downward relation **c-by**

h/H **c-by** l/L, iff

Under the given circumstances,

- the agent, **by** doing l/L, exemplifies an act h of type H;
- doing h/H **is effected / accomplished by** exemplifying an act l of type L.

A simple intuitive description of the relation between the generating act l/L and the generated act h/H derives from these definitions; it holds in both cases: Under the given circumstances, doing L is a way, or a *method*, to do H.

### 2.6.2 The Relation of C-Constitution

Rather than striving for a general formal definition of level-generation, I will apply the notion to the more concrete three types, causal, conventional, and simple. I will also introduce a different term, and with it a slightly different perspective: the notion of level-generation emphasizes the *process* of creating additional categorizations for a given act-TT. In the following I will focus rather on the conceptual *relation* between the act-TTs, and speak of “c-constitution”. Thus, the following definition of c-constitution can *mutatis mutandis* be taken as a definition of level-generation:

(17) **The relation c-const**

Let  $l/L$  and  $h/H$  be two acts such that  $l$  and  $h$  are acts by the same agent that occupy the same time, but are not co-temporal.

Under given circumstances  $c$ , an act  $l/L$  **c-constitutes**  $h/H$

$l/L \text{ c-const } h/H$ , or  $l/L \uparrow h/H$

iff one of the following two relations holds:

- $h/H \text{ c-in } l/L$  – In doing  $l/L$ , the agent exemplifies an act  $h$  of type  $H$ , or
- $h/H \text{ c-by } l/L$  – By doing  $l/L$ , the agent exemplifies an act  $h$  of type  $H$ .

## 3 Cascades and Verb Classes

In this section, I will apply the cascade approach to verb meanings, that is, lexicalized act-TTs. Goldman never did this, although, of course, he used English verbs for referring to the act-types he discussed. The recognition of the fact that Goldman's theory applies to TTs opens the way to consider level-generation as a relation between act-types, abstracting away of the particular circumstances under which a TT is exemplified. The cognitive perspective developed here allows us to apply the theory to lexical verb meanings if we assume, as I do, that these consist in event concepts that cognitively represent the type of event a verb denotes.

Applying cascade theory to lexical action verb meanings and to certain morphological and grammatical phenomena will yield ample evidence for the relevance of the approach to verb semantics. We will start out with the distinction between basic and non-basic act-TTs and demonstrate that most verbs appear to denote non-basic act-types.

### 3.1 Basic Versus Non-basic Act-Types

The notion of level-generation raises the question whether there is a basic level of action. Goldman's (1970) answer is positive. His examples of basic act-types include the following:

- (18) extending one's arm
- moving one's finger
- bending one's knee
- shrugging one's shoulder
- opening one's eyes
- turning one's head
- puckering one's lips
- wrinkling one's nose [p. 18]

Informally, a type of action is basic if it does not require a generating act of a different type in order to come about. Basic act-types are exemplified immediately, not by means of level-generation. A convenient test for non-basic act-types is to check if there are different types of act for implementing it. For example, depending on the circumstances, an electric light may be turned on by doing various more basic things, like flipping a light switch, triggering a motion detector, using a smart phone touch display, or giving a voice command to an electronic device that controls the light. Thus, 'turn on the light' is not a basic act-type. Similarly, if you are working at a computer, you may bring the cursor on the screen to a certain position by various methods, including a mouse click, using a mousepad, arrow keys on your keyboard, or touching the screen, if it is a touchscreen. Even these act-types are not basic, though; basic are just the simple bodily movements. By the way, none of the act-types displayed in the act-trees in (1) at the lowest level displayed is basic.

According to Goldman [p. 67], all action is caused by a current *want* to act correspondently. Essentially, he defines basic act-types as things an agent would do if they had the want to do so and were in standard condition with respect to this type of act, *and* if the act can be brought about without level-generation. Basicness is primarily defined for act-types, and derivatively for act-TTs.<sup>11</sup>

### 3.2 Verbs of Basic and Non-basic Action

The meaning of a verb describes a type of situation; for action verbs, it describes a type of act. The distinction between basic and non-basic act-types therefore immediately

---

<sup>11</sup> Due to Goldman's definition, basic acts are necessarily intentional. They may, however, level-generate acts that are not intended. This is an important point of the theory, but it will not play a prominent role in this paper.

**Table 1** 100 most frequent English action verbs (verbs of social action are written in italics)

01	<i>say</i>	go	make	take	come	<i>give</i>	look	use	<i>tell</i>	put
11	work	leave	<i>show</i>	<i>ask</i>	try	<i>call</i>	<i>provide</i>	keep	hold	turn
21	bring	begin	follow	<i>help</i>	write	run	set	move	play	<i>pay</i>
31	<i>meet</i>	<i>lead</i>	<i>allow</i>	carry	produce	<i>talk</i>	<i>offer</i>	consider	<i>suggest</i>	let
41	sit	continue	add	change	<i>buy</i>	<i>speak</i>	<i>send</i>	decide	<i>win</i>	<i>describe</i>
51	<i>agree</i>	build	read	reach	open	<i>spend</i>	return	draw	create	<i>sell</i>
61	cause	walk	accept	wait	pass	lie	apply	base	raise	increase
71	<i>report</i>	watch	learn	cover	<i>explain</i>	<i>claim</i>	break	support	form	cut
81	reduce	<i>establish</i>	<i>join</i>	bear	achieve	seek	deal	choose	fail	<i>serve</i>
91	<i>represent</i>	kill	drive	<i>discuss</i>	place	<i>argue</i>	<i>prove</i>	<i>introduce</i>	pick	enjoy

carries over to verbs. If one takes a look at corpus and dictionary data, it turns out that non-basicness of action verbs is the rule rather than the exception.

Table 1 displays the 100 most frequent English action verbs, among the 156 most frequent verbs in all. The table was obtained by checking the entries in the online Oxford Dictionary of English<sup>12</sup> (ODE) for the most frequent English verbs in the online British National Corpus. A verb was counted as an action verb if the first sense in the dictionary entry has an agentive, non-stative description. It was classified as non-basic if the definition was in terms of multiple synchronous or sequential action, if the method was left open, or if a cascade-like definition is given (“do ... by doing ---”). In the table, verbs of social action are marked with italics. Social action is necessarily non-basic, as its social character derives from social rules. For any type of social action, a generating physical act is required that under circumstances will count as that type of social action, according to some rule. Thus, concepts for social act-types always involve conventional generation.<sup>13</sup> I classified verbs as social if the sense description mentions interaction with other persons.

Among the one-hundred action verbs, there is not a single example of a clearly basic-act verb. One verb might be a candidate: The ODE describes the first sense of *stay* as ‘remain in the same place’<sup>14</sup>; it is a borderline case, however, and the fact that it seems basic may just be due to it not involving doing anything concrete. Certain verbs in the list may appear basic, but they aren’t. For example *say* is not basic because saying something involves a complex cascade of actions, starting from the basic acts of what we do with our articulatory organs in order to produce speech sounds; the sound productions may or may not constitute productions of linguistic sounds like vowels and consonants; even if they do, they need not necessarily constitute acts of

<sup>12</sup>Oxford Dictionary of English: <https://en.oxforddictionaries.com/>.

<sup>13</sup>See, for example, Searle (1995) on the distinction of what he calls “brute facts” and “institutional facts”. The latter form our social reality. They are what they are by social agreement. Constitutive rules of the form “X counts as Y in context C” [p. 28] create the social reality, including social action. This concept closely resembles Goldman’s notion of conventional level-generation, but Searle does not refer to Goldman’s work.

<sup>14</sup><https://en.oxforddictionaries.com/definition/stay>, accessed Jan 15 2018.

ultimately producing ordinary words and grammatical sentences. I will come back to this special case of action in the brief discussion of Austin's speech act cascade in Sect. 5.1. Even a seemingly elementary verb like *sit* is not basic (as an action verb): depending on what the agent sits on, a chair, a bike, a swing, etc. the action requires different physical activities; *sit* may also mean 'sit up' from a lying position, or 'sit down'—asking for yet different physical action. Apart from these senses, there is the transitive use of *sit* as in *sit the child on one's shoulder*. Even if certain verbs denote action that is closely related to a particular body part, like *kick*, they are not necessarily basic, as one can, for example, kick with various parts of the foot, with one's shin, one's knee or thigh—variants of kicking that are executed by different more basic types of action.

As a result, it appears that there may be no basic-act verbs at all among the 100 most frequent English verbs. Are there any basic-act verbs in English, verbs that invariably denote basic action rather than what is accomplished by some type of more basic action? The verbs in Goldman's basic action examples in (18)—*extend, move, bend, shrug, open, turn, pucker, wrinkle*—are not in themselves verbs of basic action. In Goldman's examples, they are all transitive verbs and their basicness depends on the choice of a particular body-part as the object argument. For types of object other than one's own body-parts ('move the table', 'turn the pancake', 'open the door'), there would be various methods of enactment available. Some of the verbs have intransitive action uses—*move, bend, shrug, and turn*; among them, *shrug* is a candidate for a basic-action verb because to shrug is the same as to shrug one's shoulder; maybe intransitive *bend* is another one.

It is not surprising that there are so few verbs that denote basic acts. The vocabulary of natural language serves communication in, and about, *our reality*, and this is to a large part social reality. Verbs of action are used in order to describe what people do. If we were restricted to verbs of basic action, it would be extremely hard, if not impossible, to describe what people are really doing (try to say that you are writing an article by reporting the basic physical movements you make to do so—no-one would understand what you are describing). Quite generally, it seems, we communicate about what people do on considerably advanced levels of cascading. Verbs like *help* supply a good illustration of the 'abstractness' of action concepts. Ranking 24 in the above list, it is central vocabulary. According to the analysis in Engelberg (2005), the verb means essentially 'do something for somebody that improves their situation'. The concept of helping leaves open what the generating action would be concretely; in fact, an action of almost any type may constitute help in one situation, and the contrary in another, and the very same act-token may constitute help for one person and a big problem for another. In social life, improving others' situation is of utmost importance; it applies to all kinds of situation in our complex lives; we *need* general verbs like this.

For another source on basicness or nonbasicness, one may take a look at Levin's (1993) *English Verb Classes and Alternations*, where a comprehensive collection of semantic verb classes is compiled and described. There are 49 major classes distinguished, almost all of them action verbs—not a single class is basic-action.

### 3.3 Criterion Predicates

Goldman's theory of action was not really taken up in semantic theories of verb meaning.<sup>15</sup> There is, though, a small thread of discussion on the semantic analysis of *by* gerunds where a two-level view on the meaning of selected types of action verb is adopted. The discussion starts out with Kearns (2003). Kearns distinguishes two special classes of action predicates which she dubs “causative upshots” and “criterion predicates”. Causative upshots are transitive predicates like *cure the patient* or *convince s.o.* [p. 599]; they denote the achievement of some sort of change by doing something more concrete, e.g. curing someone by administering a certain treatment, or convincing someone by presenting evidence. Criterion predicates are often intransitive and not inherently causative; they include predicates such as *make a mistake*, *break the law*, *score a goal*, or *prove a theorem*. As with *help*, the predicate requires that something be done that fulfils a given criterion, while the method is left open; it can be specified with a *by* or *in* locution (recall the example in (13)). For both types of predicate there is, in Kearns' terms, a “host” and a “parasite” [pp. 600–1]. The “more abstract” parasite, the causative upshot or criterion predicate, is denoted by the verb and is implemented, or accomplished, by the “more concrete” host. For example, the parasite is ‘breaking-the-law’ and the host is a theft; the parasite is ‘curing-the-patient’ and the host is administering the treatment. Clearly, Kearns' hosts level-generates the parasites. Kearns does not mention Goldman's work, though. Her analyses are confined to two levels, and to two special classes of generated act-types.

The two classes of verbs were taken up in Sæbø (2008, 2016). He chooses different terms for Kearns' causative upshots (“manner-neutral causatives” in 2008, “method-neutral causatives” in 2016); hosts and parasites he calls concrete and abstract.

Notably, the “hosts”, or more concrete acts, are not basic in the sense explained here, at least not necessarily so; they may be high-level act-types. What matters here, is that the two authors distinguish within one verb meaning different levels of action related by, in fact, level-generation.

### 3.4 Means of Explicit Level-Generation

In addition to this lexical evidence for cascade-structure action concepts, there are numerous lexical and grammatical mechanisms operating on verbs and their lexical meanings to the effect of generating further cascade levels. Some of them involve word formation, for example affixation, or conversion from a different word class, others employ certain grammatical constructions, or types of adverbial. The examples in the following are chosen for the sake of illustration; they do not provide a systematic

---

<sup>15</sup>The theory was taken up and developed further in Clark's (1996) theory of communication where he introduces the concept of “action ladder”. However, Clark did not apply the notion of level-generation to verb semantics.

survey, but represent just the tip of an iceberg. Almost all the cases described involve augmentation along with level-generation. The augmentation of the underlying action concept iconically corresponds to the augmentation by word formation and/or syntax at expression level.

### 3.4.1 Adding a Level of Social Interaction

Many lexical and grammatical processes add a further argument<sup>16</sup> to a given action concept. This amounts to augmentation of the underlying concept, but in addition c-constitution is involved, on top of the augmentation. I will discuss the addition of arguments of the type ‘person’; this will inevitably have the effect of cascading to a level of social interaction.

Many basic types of bodily action are used as non-verbal signals in communication. For example, the verb expressions *smile*, *frown*, *raise one’s brows*, *wink*, *nod*, *shrug*, *bow*, *kneel down*, *fold one’s hands*, *scratch one’s head*, *wave one’s hand*, and others can also denote communicative action. They do so invariably if they are used with a prepositional phrase that adds an addressee: ‘smile/wink/wave/frown at someone’. German has verb prefixes such as in *zu-zwinkern* (‘wink at’) or *an-lächeln* (‘smile at’) which serve the same effect of enriching the argument structure with an addressee.<sup>17</sup> (19a) is an example that attests the social-level relevance of *zuzwinkern*. The concept of *zuzwinkern* has the informal cascade structure in (19b).

- (19) a. *Mein Lieber, wenn du nicht verheiratet wärst, dann könnte ich dir jetzt zuzwinkern.*  
[DWDS]  
‘My dear, if you were not married, I could now wink at you.’  
b. Cascade: ‘zuzwinkern’: ‘zwinkern’ ⊏ ‘zwinkern’ + *addressee* ↑ ‘zuzwinkern’

German *an* and *zu* can also be used as prepositions marking an additional addressee argument for verbs of communication: *schreiben an* + accusative NP ‘write to’ or *sprechen zu* + dative NP ‘speak to’.

Similar to these cases are applicative constructions (Van Valin and LaPolla 1997: 337–8). Japanese has several such constructions consisting of two verbs. The first verb is in the gerund *-te* form and the second a verb of possession transfer, such as *ageru* ‘give upward’ and *kureru* ‘give downward’; the direction component is metaphorically used for expressing ‘give to superior’ or ‘give to inferior’. A speaker will always treat the addressee as socially superior and themselves as inferior; therefore the beneficiary in the *-te ageru* construction will typically be the other, and the agent typically the self or someone related to the self. The complex expression is used to describe doing a favor.<sup>18</sup> The cascade analysis has the first verb as the generator.

<sup>16</sup>It is not relevant in this context to distinguish syntactically between complements and adjuncts; we will talk of ‘arguments’ in both cases.

<sup>17</sup>See Stiebels (1996: 163f) on the prefix *an-*.

<sup>18</sup>Martin (1975: 597–601).

- (20) a. Japanese

<i>mado</i>	<i>o</i>	<i>ake-</i>	<i>te</i>	<i>age</i>	<i>ta</i>
window	ACCUSATIVE	open-	GERUND	give-	PAST
‘I opened the window for you/her/him/them’					

- b. Cascade:

‘open the window’ ⊓ ‘open the window’ + *superior addressee*  
 ↑ ‘do superior a favor’

Thus, the construction has the structure of a criterion predicate, with the method specified. A similar construction in Mandarin is discussed in Tsai (2012). It makes use of the verb *gěi* 给 ‘give’ that is otherwise also used as a standard verb of giving (Chang 2016: 251–2).<sup>19</sup>

- (21) a. Mandarin (Tsai 2012, p. 5)

<i>gei</i>	<i>wo</i>	<i>gui-</i>	<i>xia!</i>
AFF	me	kneel-	down
‘Kneel down for my sake!’			

Van Valin and LaPolla (1997, p. 384) describe beneficiary constructions in Lakhota with essentially the same semantics. German has a special use of the dative in such cases<sup>20</sup>:

- (22) German

<i>Er</i>	<i>hat</i>	<i>ihr</i>	<i>die</i>	<i>Tür</i>	<i>aufgehalten</i>
he	has	her.DATIVE	the	door	kept open
‘he [has] kept the door open for her’					

As witnessed by the translation, English has a *for*-complement construction with the same function.

### 3.4.2 Adding a Level of Achieving a Result

Predicate expressions such as *hammer flat* or *drink empty* consist of a verb of action and a predicative adjective that denotes a resulting state of the object acted upon. Resultatives of this type denote an action that is generated by an act of the type of the base verb; for example, *hammer flat* denotes a cascade of the structure ‘hammer ...’ ↑ ‘flatten’, and *drink empty* a cascade ‘drink ...’ ↑ ‘empty<sub>verb</sub>’. However, the cascade first requires an augmentation that adds the affected object. Thus, the analysis again requires two cascade steps:

---

<sup>19</sup>Tone diacritics are not given in the source.

<sup>20</sup>Wegener (1985: 94–6) on *dativus commodi*.

- (23) a. ‘hammer’  $\sqsubset$  ‘hammer’ + ‘on x’  $\uparrow$  ‘flatten x’  
 b. ‘drink’  $\sqsubset$  ‘drink’ + ‘from x’  $\uparrow$  ‘empty x’

Dowty (1979), and many others since, analyzed this type of construction as causative in the sense that, for example, *drink the glass empty* means ‘drink from the glass and [thereby] cause the glass to become empty’ (Dowty 1979: 93). This is reflected by the analysis in (23) if  $\uparrow$  is taken as representing the causal type of level-generation. German has a lot of particle verbs with a resultative particle such as *tot-* ‘dead’ in *tot-schießen* ‘shoot to death’, *klein-* ‘small, little’ in *kleinschneiden* ‘cut into small pieces, chip’ or *an-* ‘on’ in *anknipsen* ‘to flick on’; these can be analysed analogously.

Van Valin and LaPolla (1997: 90) mention verbs of killing in Lakhota; they have the form of compounds with the first part indicating the method of killing, and the second a verb *t'a* that means ‘dead / to die’, for example *ka-t'a* ‘strike to death’ (*ka-* ‘by striking’), *ya-t'a* ‘bite to death’ (*ya-* ‘with the teeth’), *yu-t'a* ‘strangle’ (*yu-* ‘with the hands’). English can generally use the addition *to death* for level-generating a predicate of killing. German has a series of verbs of killing with the prefix *er-* that does not have much of a lexical meaning on its own, but rather constructional meaning in this type of verb formation: *erschießen* (‘shoot to death’), *erschlagen* (‘beat to death’), *erwürgen* (‘choke/strangle to death’), *erhängen* (‘hang’), *erdücken* (‘crush to death’), and several more.<sup>21</sup>—The generating act-type fails to be specified in cases of conversion of adjectives to verbs; the adjective denotes the resulting state of the object of an unspecified action: *empty, fill, smooth*, etc. These verbs are method-neutral predicates in the sense of Sæbø (2016).

### 3.4.3 Adding a Level of Appraisal

A further type of cascade extension adds an appraisal to the action-verb concept. German has a productive word formation pattern that derives from almost arbitrary verbs of action a verb used to express failure; these verbs have been dubbed ‘erratic’ verbs (see Fleischhauer 2016: 293). One variant of the derivation adds the prefix *ver-* to a transitive verb and yields another transitive verb (*die Hecke verschneiden*, ‘cut the hedge in the wrong way’<sup>22</sup>); a second type adds the same prefix and the verb is reflexivized as to form an intransitive predication (*sich verschneiden* ‘cut in the wrong way’). This derivation adds a cascade level of failure: ‘cut’  $\uparrow$  ‘fail’. Thus, this is another mechanism that produces criterion predicates. The highest level of the cascade is fairly unspecific, but the cascade as a whole yields the meaning expressed. English has some erratic verbs with the prefix *mis-*: *misunderstand, misdirect, mishear*, but the pattern is far less productive than the German one.<sup>23</sup>

<sup>21</sup>Stiebels (1996: 234–5).

<sup>22</sup>Stiebels’ example in her discussion of this *ver-* derivation (1996: 143–51).

<sup>23</sup>Goldman (1970: 17) mentions erratic ‘misspeak’, ‘miscalculate’, and ‘miscount’ as examples of act-types that “preclude intentionality”. While the underlying basic act is intentional, it happens to generate unintended action.

Other constructions across languages serve the generation of a level of ‘doing too much’: cf. English *overcook*, *overheat*, *overpay* etc.; Russian uses the prefix *pere-* in a similar way (*pere-gret'* ‘overheat’).<sup>24</sup> Japanese has verb compounds with the second verb *-sugi-ru* ‘exceed’, for example *nomi-('drink')-sugi-ru* ‘drink too much’.<sup>25</sup>

A two-verb construction in Mandarin with the second verb 玩 *wán* ‘play’ can be used to express the level-generation of acting for pleasure:

- (24) Mandarin (Liu Fan, from the BCC corpus)

*Wǒ xiàwǔ chūqù hé péngyou guàngjiē wán ne*  
I afternoon go.out with friend go.shopping play PRT  
'I go out to shopping with my friend for fun.'

German has a very productive adverb formation that adds *-erweise* to an adjective or a present participle stem. This type of adverb is used for evaluating an act, or more generally an event or a state. Examples include *dummerweise* ‘stupidly’, *erstaunlicherweise* ‘surprisingly’, *unnötigerweise* ('unnecessarily'), *glücklicherweise* ('luckily'), and hundreds more. They correspond to English adverbs in sentence-initial use.

- (25) German (DWDS corpus)

*Dummerweise hatten wir keine Schneemäntel angezogen.*  
'Stupidly, we hadn't put on snow coats.'

This type of adverb projects the verb to a criterion-predication level. For example, adding *dummerweise* to a verb V, has the effect of [V] ↑ ‘do something stupid’.

### 3.5 Implicit Level-Generation

It may be worthwhile considering cases of “integrated” augmentation generation of the types discussed above as they provide a glimpse into the decompositional structure of certain types of action concept.

**Appraisal.** One group with an integrated specific evaluation is constituted by verbs of forbidden action, e.g. *lie*, *steal*, *trespass*, *rob*, *rape*, *murder*, and many others. These add to the concept of a particular type of action a level ‘do something forbidden/illegal’. Thus, there is a cascade relationship between ‘kill’ and ‘murder’. ‘Murder’ can project further to ‘assassinate’ if the victim is an important person, giving rise to elaborate cascades such as ‘shoot’ ⊓ ‘shoot at y’ ↑ ‘kill y’ ↑ ‘murder y’ ↑ ‘assassinate y’.

**Result.** Van Valin and LaPolla (1997) distinguish causative and active accomplishments, and achievements. Causative accomplishments are verbs like *kill*: the agent

<sup>24</sup>See Zinova (2016: 146–51) on a frame analysis of the meanings of *pere-*.

<sup>25</sup>Martin (1975: 434–8) on the “excessive” construction.

does something that causes somebody to die. The authors apply the following general half-formal analysis to this type of action verb [pp. 188–9].<sup>26</sup>

- (26) [do x, [**predicate**<sub>1</sub>(x, (y))]] CAUSE [BECOME **predicate**<sub>2</sub>(x) or (y)]].

This reads essentially as follows: agent x does something of the type **predicate**<sub>1</sub> which causes x or y to change into the condition denoted by **predicate**<sub>2</sub>. The first part of the analysis—do x, [**predicate**<sub>1</sub>(x, (y)))—describes an action by the agent x (that possibly involves another participant y); according to the second part—CAUSE [BECOME **predicate**<sub>2</sub>(x) or (y)]—x's doing causes x or y to enter the condition described by the second predicate. The whole formula describes the constitutive condition for causal generation<sup>27</sup>:

- (27) **predicate**<sub>1</sub>(x, (y)) ↑ [ x MAKE [BECOME **predicate**<sub>2</sub>(x) or (y)]]]

Causative achievement and accomplishment verbs with an agent argument are abundant in natural languages. Typically, the generating level of the more basic method action is not specified.

**Signaling.** As mentioned above, some action verbs of basic or near-basic level can be used to denote a social-level act of signaling (*smile, frown, harrumph, nod, shrug*, and others). If used in this sense, they incorporate generation of a social level. As social agents, equipped with the “sense-making machines” our minds are, we usually try to come up with a construal of the acts of others as meaningful beyond the mere act. The verbs mentioned reflect this tendency by incorporating a higher cascade level in lexicalized meaning variants.

## 4 Cascades and Frames

Application of Goldman's approach to psychology calls for a framework for modelling cognitive representations. I apply the theory of Barsalou frames as further developed in the Düsseldorf context of research on the structure of representations.<sup>28</sup> The framework is applied to the decompositional analysis of lexical meanings and

<sup>26</sup>The analysis goes back to Dowty (1979), who relates to McCawley (1968) for the structure of the analysis.

<sup>27</sup>In the Dowty formula in (26), ‘CAUSE’ denotes a relation between events: the event denoted by the first predicate causes the event denoted by the second. In Goldman's definition of causal generation in (4.1.b), ‘cause’ is used as an agentive verb: the agent causes an event e. The two uses of ‘cause’ correspond to two senses of the verb *cause*. In order to distinguish between these two senses, I used MAKE for agentive causation in (27). I am grateful to Wilhelm Geuder and Ekaterina Gabrovska for making me aware of this point.

<sup>28</sup>The Collaborative Research Center 991 on “The structure of representations in language, cognition and science”. For representative work on this approach see Petersen (2007), Kallmeyer and Osswald (2013), Löbner (2014, 2017).

the modelling of compositional processes, among other things.<sup>29</sup> I will characterize it here very briefly and then propose an integration of cascade structures into the theory.

## 4.1 Barsalou Frames

As a working hypothesis, I adopt Barsalou's Frame Hypothesis, according to which Barsalou frames constitute the universal format of concept representation in human cognition.<sup>30</sup> It is assumed that lexical meanings are concepts stored in long-term memory and that compositional meanings are concepts formed as the result of syntactic and semantic processing, essentially by unification.

According to Löbner's (2017) formal theory of Barsalou frames, a *frame structure* is a coherent network of nodes connected by functional attributes. The nodes represent individuals in a global universe of discourse. The attributes are functions that for individuals of an appropriate type return another individual of the same or another type as value. For example, the attribute SIZE returns the individual size for all individuals that have size; the attribute MOTHER returns the mother for every animal with parents; the attribute HEAD returns the head for those things that have a head. The values of attributes may carry their own attributes; thus, frame structures are recursive. In a frame, type restrictions may be imposed on the nodes, that is, conditions specifying that the entity represented by the node belong to a certain subset of the universe. The frame structures defined in Löbner (2017) are first-order in that the underlying ontology provides a universe of discourse, the set of all individuals, and the attributes are functions that return individuals to individuals. The universe does not contain second-order entities such as properties, relations, attributes, or first-order frames. Frame structures can be translated into an appropriate first-order predicate logic language (see Löbner 2017: 99–109 for details).

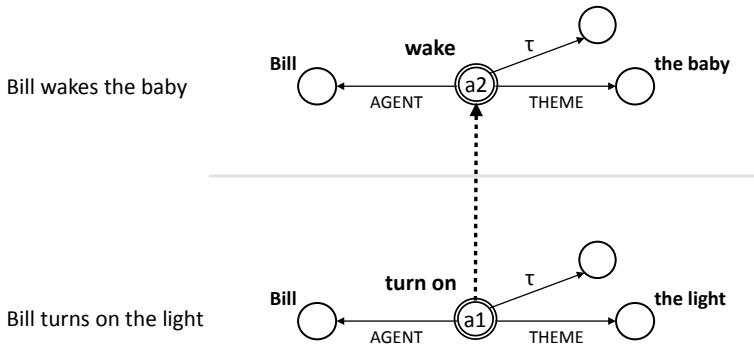
Frames are usually represented by frame diagrams (see examples below), or else by attribute value matrices. I will use diagrams. There is always a distinguished central node that represents the individual described by the whole frame. Frames have the same double nature as Goldmanian act-TTs: they represent a token of a type. A frame diagram as a whole provides a type description of the token represented by the central node; the analogue holds for frames represented by attribute-value matrices.

In the context here, we exclusively deal with frames for actions. Actions are a particular type of individual in the universe, a subtype of events. All events have an attribute  $\tau$  for the time they occupy; therefore every action frame has this attribute on the central act node. Actions have an agent whence the act node in an action frame carries an attribute AGENT. For the current discussion in the context of a theory of human action, it will be assumed that agents are persons. An action frame may

---

<sup>29</sup>See, for example, the contributions by Andreou and Petitjean, Balogh and Osswald, and Gamerschlag and Petersen in this volume.

<sup>30</sup>See Barsalou (1992: 21) for the original source, and Löbner (2014) for its application to language.



**Fig. 2** Cascade formed by two frames

contain more attributes of the act, corresponding to more semantic roles such as THEME, PATIENT, INSTRUMENT, GOAL etc.<sup>31</sup>

## 4.2 Cascades in Frame Theory

The question arises if cascades are another variant of frames. Löbner (2017) allows only first-order attributes in frames. The cascade relations c-constitution, c-in, c-by, and subsumption, however, are essentially and irreducibly second-order, because they relate types, i.e. whole first-order frames. Apart from that, the upward relations are not functions. Due to transitivity, a level-generating act-token does not project to a uniquely defined token it generates. In addition, level-generation may branch upwards. Thus the cascade relations cannot figure as attributes *within* first-order frames. I will integrate them into frame theory as second-order relations *between* first-order frames.

Let us consider a simple two-level cascade for illustrating the interplay of frame representation and c-constitution:

(28) a1/'Bill turns on the light' **c-const** a2/'Bill wakes the baby'

The cascade diagram in Fig. 2 contains the frames for a1/'Bill turns on the light' and for a2/'Bill wakes the baby' at the lower and the upper level, respectively. The two frames are parallel in structure. They have a central act node that represents an act of the type indicated by the bold-face type label. In both frames, the action nodes

<sup>31</sup>For more elaborate verb frames, see for example Kallmeyer and Osswald (2013), Naumann (2013), Gamerschlag et al. (2014), Löbner (2017), and the contributions to this volume mentioned. Verb frames that only display attributes for semantic roles and the time  $\tau$  are a gross simplification of what the decomposition of lexical verb meanings ultimately calls for. However, one is always free to reduce frame representations to what is needed in the context of discussion. For the needs of this paper, case frames will suffice.

carry the attributes AGENT and  $\tau$ . Both frames also have a THEME attribute on the central node, of different nature. As the two frames are related by c-constitution, the attributes AGENT and  $\tau$  necessarily both take the same value in the lower and the upper frame. The identity of agent and time cannot be expressed by linking the attributes in both frames to one value node; attributes cannot take values in another frame than their argument node belongs to. The identity of values can only be accomplished by assigning the same individuals as values for the two attributes, respectively. The dashed upward arrow in Fig. 2 stands for the relation of c-constitution between the two acts.

A structure formed by more than one first-order frame is itself second-order, that is, a hyperframe. Hyperframe structures are a natural extension of first-order frame theory. For example, if one is to model scripts with frames, one will have to design hyperframes that consist of first-order action frames for subsequent acts, connected in an appropriate way.

## 5 The Writing Cascade

We will now turn to an elaborate example, the cascade for the act-type ‘write by hand’. It will be used to discuss the consequences that the adoption of the cascade model to lexical verb meanings has for semantic theory. As a prelude, we will have a brief look at Austin’s (1962) speech act model. Austin’s analysis anticipated Goldman’s multilevel theory of action; Goldman mentions it as such in his introduction [p. 8].<sup>32</sup> The speech act cascade also prepares the discussion of the writing cascade in the section to follow because the upper levels of the speech act cascade also appear in the write [act] cascade.

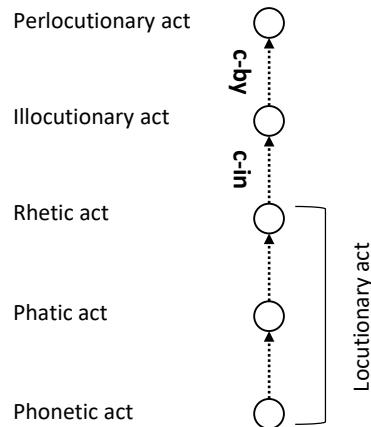
### 5.1 Austin’s Speech Act Cascade

Austin’s (1962) analysis of speech acts constitutes a classical example of a cascade. Austin distinguishes five levels of action in an ordinary verbal utterance (Fig. 3). The “locutionary” level consists in saying something with a particular sense and reference in the given context of utterance. Within the locutionary act, Austin makes a finer distinction into three levels: with the “phonetic act”, the speaker produces speech sounds; the “phatic act” is “the uttering of certain vocables or words, that is, noises of certain types, belonging to and as belonging to, a certain vocabulary, conforming to and as conforming to a certain grammar.” (Austin 1962: 95); the “rhetic act” is “the performance of an act of using those vocables with a certain more-or-less definite

---

<sup>32</sup> A recent work that links Austin’s speech act model to Goldman’s level-generation is Moltmann (2017). She applies the level approach in particular to the distinction of locutionary and illocutionary act.

**Fig. 3** Austin's speech act cascade

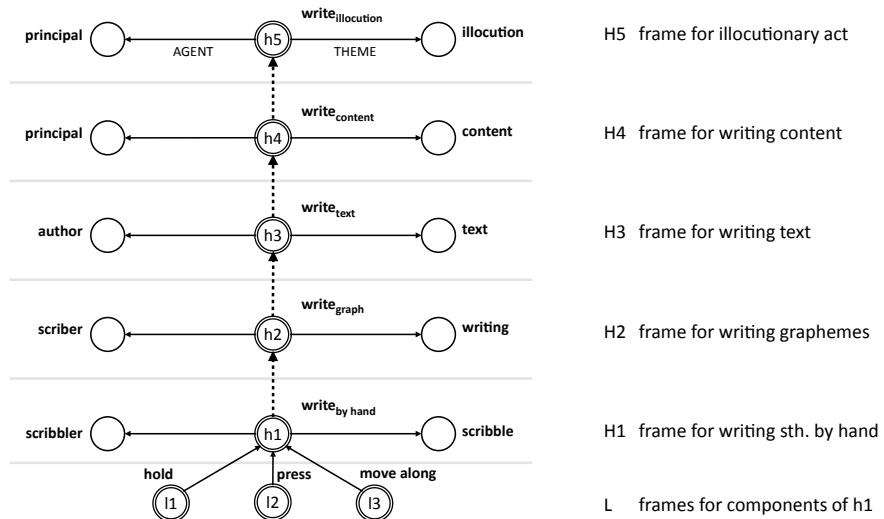


sense and reference.” [p. 95]. The phonetic act generates the phatic act, and this in turn the rhetic act. Austin continues [p. 98], “To perform a locutionary act is in general, we may say, also and *eo ipso* to perform an *illocutionary act*”. Austin calls this level the *illocutionary act* in order to emphasize that it is done *in* performing the locutionary act. He thus explicitly assumes a *c-in* relation between illocution and locution. The achievement of the illocutionary act—a promise, an answer to a question, etc.—only succeeds if complex “felicity conditions” [pp. 25–38] are fulfilled. Austin discussed these conditions in detail, thereby offering an elaborate case study of the “circumstances” involved in these cases of level-generation.

Finally, by performing an illocutionary act, the speaker may execute a “perlocutionary act” that consists in causing a particular effect, for example, convincing, offending, or delighting the addressee. Austin calls it *perlocution* because it is done *by* performing the illocution [p. 108]. “[T]he perlocutionary act always includes some consequences” [p. 107]. Unlike the lower four levels of a speech act, the perlocutionary act may or may not be intended. The nature of the four level-generations is a combination of conventional and simple for phatic, rhetic, and illocutionary act; the level-generation of the perlocutionary act from the illocutionary act is causal; it does not involve convention [p. 121].

## 5.2 The Cascade Structure of Writing by Hand

We will now proceed to an example that is suitable to illustrate and discuss central aspects of applying the cascade approach to verb semantics. Figure 4 displays a cascade for the concept of writing by hand. This concept essentially constitutes the lexical meaning of the verb (except for the specification of the lowest level which we will argue in Sect. 6.1 is not specified in the lexical entry). It is roughly analogous to Austin's cascade, but I will elaborate it more, commenting on the single-level frames



**Fig. 4** The cascade for writing by hand

and their relationships. The writing cascade has a lowest level of three co-temporal acts: the agent holds a writing implement in their hand, presses its writing part on some surface, and moves it along leaving a visible trace. Compound augmentation integrates the three co-temporal acts into the act-type at H1 ‘write by hand’, the first level that can be called writing, in the sense of producing visible lines and shapes. For reasons of space, the three frames for the acts of holding, pressing, and moving along are only represented by their central act nodes. In fact, they share the agent and the action time among them; they also have the same theme argument (i.e. the pen or other writing implement); the acts of pressing and moving share the surface as a third argument. Actually, the process of handwriting is even more complex; usually, the pen will not be in continuous contact with the surface since writing will require to lift the pen and move it to a different position on the surface. We neglect this aspect here.

The higher Levels H1 to H5 consist of action frames that each have an AGENT and a PRODUCT attribute (the attribute arrows are labeled accordingly only in the highest level). If Level H1 produces perceptible forms of writing on the surface, it generates Level H2 ‘write<sub>graph</sub>’ of producing graphemes. Graphemes, in turn, may or may not constitute linguistic text: under circumstances, Level H2 generates Level H3 ‘write<sub>text</sub>’. Again under circumstances, writing text constitutes a fourth Level H4 ‘write<sub>content</sub>’. Writing verbal content corresponds to the locutionary level in Austin’s cascade. To this level adds an illocutionary level H5 ‘write<sub>illocution</sub>’, for example, an application, an excuse, a reply, a request, etc. The specific type labels for the agents will be explained in Sect. 5.4. A perlocutionary level is not assumed to figure in the concept of writing.

At each cascade level, the act is embedded in a different context, and each context comes with different conditions and requirements. The context of Level H1 is the same as, for example, the context of a drawing activity. The agent needs a surface such as a sheet of paper and a pen or other implement, maybe along with ink, paint, etc. The agent needs to be able to hold the implement and move it along on the surface at some level of motor control. The agent determines readability in terms of the size of writing, the visibility of the writing material on the surface, the durability of the product; they may be concerned with highlighting parts of the writing by different color or style. The product at Level H1 can be copied or scanned; if properly processed, it can be stored on an electronic device. At Level H2, the agent bothers about a writing system and a writing style; they need to command the skill of writing; they will write legibly or not. The Level 3 agent is concerned with choosing a language, with orthography and grammar; they need be in sufficient command of the language. At Level H4, the agent is an author of content, whereby the agent potentially relates to other content and its authors; for larger texts, the author is concerned with aspects such as coherence and structure which are crucial for comprehensibility. Obviously, producing text involves more abilities than just knowing the language. It is at the illocutionary level H5 that the agent enters social interaction with a reader addressee, possibly initiating or continuing a sequential exchange; the agent at this level will choose an appropriate type of text, a style and a tone of expression, which requires the relevant social skills. At each level, different criteria of successful action obtain. And each level is motivated and informed by what it serves to level-generate.

### 5.3 Types of Products and Levels of Manner Modification

Depending on the level, writing brings about different types of product, for example, lines, letters and characters, words, coherent text, illocutions, etc. This amounts to different selectional restrictions for each level. Correspondingly, if the verb *write* is complemented with a direct object such as *whorls*, *e's*, “*mama*”, “*I'm to the cafeteria*”, *a receipt*, etc., an appropriate level within the cascade will be selected for application. If one were to describe the selectional restrictions for the theme argument of *write* in a single-level approach, one would run into an inconsistent type assignment for the product argument.<sup>33</sup>

---

<sup>33</sup>One approach that deals with this problem is the assumption of “dot objects” (see for example Pustejovsky 2009; Asher 2011). Dot objects are of a composite type, such as *physical\_object* • *information* for ‘book’. There is a vague connection between this approach and cascade theory, if the notion of cascade is extended to objects (see below), but the relationship is too unclear to be addressed here. The dot-objects approach raises many questions: What is the ontological character of dot objects—are they one object or more? Which types of object can be combined to form dot objects? What is the relationship between the elements and the whole? At present, I can state about that much: there are cases of dot objects that form a cascade, in particular, dot objects of the type *action* • *action* as discussed in Bücking (2014). There are other cases that might constitute cascades if the notion is generalized as to also cover objects. But there are also cases that clearly do not form

The level-distinction is equally relevant for the analysis of manner modification. (29) lists manner modifiers of *write* that are level-specific; others like *slowly* or *beautifully* may apply at more than one level.

- (29) Level-specific manner modifiers of *write* and the cascade-levels they relate to
- H1 *swiftly, shakily*
  - H2 *small, illegibly*
  - H3 *ungrammatically, in Dutch*
  - H4 *coherently, consistently, incomprehensibly, redundantly, laconically*
  - H5 *urgently, rudely*

Without requiring disambiguation or coercion, the verb combines with any-level modifiers or product specifications. Simultaneous relation to different levels is possible, such as in the following example:

- (30) *She used to write her private letters [H4] with two fingers [L] on her typewriter [L].*

## 5.4 Agencies at Cascade Levels

In Goldman's theory, the agents of the acts in a cascade are presupposed to be the same. They are, however, in different *roles*, a fact that is blurred if one uses the same generalized attribute AGENT through all levels as I did in Fig. 2 and the writing cascade; the difference becomes transparent if one uses instead the more specific role attributes that actually apply. These are in the case of writing by hand:

- (31) Level Agent's role
- L the one who holds the writing implement in hand  
the one who presses its writing part on the surface  
the one who moves it along on the surface
  - H1 the scribbler
  - H2 the scribe
  - H3 the author of the text
  - H4 the principal of the content
  - H5 the performer of a written illocutionary act

Goffman (1979) introduced the notion of “footing” in order to distinguish different roles that the participants in a verbal communication can take on.<sup>34</sup> There are producer

---

cascades, namely those like *plant<sub>N</sub> • drink<sub>N</sub>* (for ‘coffee’), of the type *source • product*, where the two objects united in the dot type do not temporally coexist. Other cases such as *producer • product* ('Honda') or *institution • printed copy* ('newspaper') are plain metonymies not requiring a special ontology. (For a treatment of metonymy in Frame Theory see Löbner 2013: 313 ff.).

<sup>34</sup>See Levinson (1988) for discussion of Goffman's notion from a linguistic point of view.

footings and recipient footings. On the producer's side, which matters here, Goffman distinguishes the roles of "principal", "author", and "animator". The principal is the one on whose behalf an utterance is made, the one who is responsible. The author chooses the words, the animator produces the verbal signals. In everyday communication, the three roles are usually enacted by the same person. In institutional settings, however, like press conferences, public speeches, court trials, examinations, and countless others, the producer footings may be distributed among more than one person, present or absent; ghostwriters choose the words they don't utter themselves, attorneys speak on behalf of their clients, typists type words not their own. In the diagram of the writing cascade in Fig. 4, the agent nodes are labeled according to Goffman's distinctions. Agentship can in principle be delegated down the cascade if the higher-level agent is in a social position to do so. A lower-level agent is responsible to their higher-level delegators; ultimately, the principal will be held responsible for the performance of all the agents involved at the lower levels.

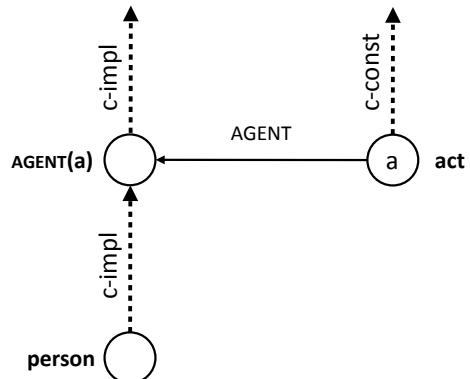
These considerations suggest a generalization of level-generation that allows for delegation of agency down the cascade, instead of strict identity of agents. In the realm of social interaction, delegated agency is a common phenomenon. For example, I may help somebody by delegating helpful action to a third party; I may pay a debt by having a third person pay who owes me money; I may break the law by making my subordinates do something illegal, and so on.

If agency does not split, there is a relation more specific than physical identity between the agent roles at the different levels—if these agents are not considered just persons but persons-in-a-particular-role. Let us assume that Erica holds a pen and moves it along a piece of paper. As such she is already in three roles, implementing the penholder, the one who presses the pen upon the paper, and the one who moves it along on the paper. If she produces script, she thereby implements a 'writer-by-hand'. The implementation cascades upwards if Erica is successful in writing graphemes, thereby producing text, content, an illocution. Under the circumstances required, the agent at a given generator level *implements* the agent at the generated higher level. As the implementation is successful only under circumstances, I will talk of "c-implementation".

The implementation relation is asymmetric: the writer-of-text implements a writer-of-content, but not vice versa, since text need not have content. It is also irreflexive: no role implements itself. And implementation is transitive. Thus, the c-implementation relation has essentially the same properties as c-constitution, except for the fact that it is a relation between persons and the roles they implement, rather than between acts. In analogy to c-constitution, I consider c-implementation as a relation between TTs, in this case persons under a particular role description, for example  $\text{Erica/AGENT(h1/write}_{\text{by hand}}\text{)}$ , that is, "Erica in the role of the agent of an act h1 of the type 'write}\_{\text{by hand}'".

C-implementation shares with c-constitution the question of grounding. Although c-implementation goes hand in hand with c-constitution of acts, the grounding of c-implementation is not just derivative from the grounding of c-constitution. Rather, for any level of action, including the basic level, taking the agent role means implementing it, for the person who acts. Hence, if I/L is the basic act-TT in a cascade to

**Fig. 5** The two levels of implementing an agent role



perform, the c-implementation chain starts with an additional prior step, taking the form in (32a), while the corresponding act-cascade is as in (32b):

- (32) a. person x **c-impl** x/AGENT(1/L) **c-impl** x/AGENT(h1/H1) **c-impl** ...  
     b. \_\_\_\_\_ l/L                           **c-const** h1/H1                   **c-const** ...

Figure 5 displays the two levels involved with agency: the person who implements the agent and the person in the agent role for a specific act. The act level may cascade further upwards.

We may assume that a person is implemented by a living human, the human by an organism, the organism by biomass, and so on. This assumption would be in line with theories that model social entities such as persons as supervenient on biological entities, and these on chemical entities, etc. The problem of grounding persons is an ontological problem of its own.

This mismatch notwithstanding, we may consider to generalize the term *c-constitution* as to also cover the c-implementation relation. It makes sense to extend the use of the term in this way: the writer-by-hand under circumstances *constitutes* a writer of graphemes, who in turn may *constitute* an author of text, and so on.

## 5.5 Objects at Cascade Levels

Goldman's notion of level-generation does not impose conditions on arguments other than agents. In view of the writing cascade, we see that it would be inadequate to assume identity of the products across levels because they exemplify ontologically different types of object. Extracting the product track from the cascade yields a multilevel conceptual description of the product on its own. The products are things of a quality that originates at Level H1, H2, etc. respectively. Again, there is a relation of constituency: under circumstances, the graphemes constitute text, the text constitutes content, the content an illocution.

The difference of description that applies to the products of writing at the levels distinguished is particularly conspicuous. This will always be the case for object arguments in action cascades of creating, destroying, or changing things, like *bake*, *break*, or *repair*. However, objects in any cascade will be in different roles, too, analogous to the agents in a cascade. Consider the following cascade, imagining circumstances that would support its formation:

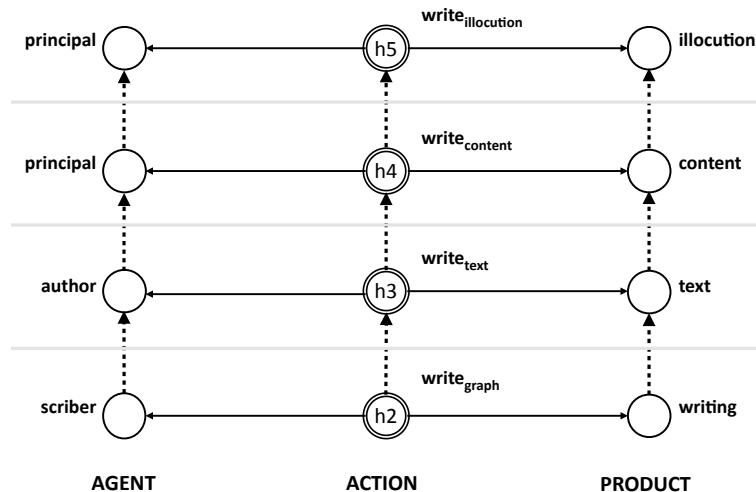
- (33) L Amy presses the power button on the TV remote control
  - H1 Amy turns on the TV
  - H2 Amy turns on the evening news
  - H3 Amy starts her daily evening TV ritual
  - H4 Amy breaks off the on-going conversation with her friend
  - H5 Amy annoys her friend

And now consider the role of the TV set at the different levels:

- (34) L The TV is a remote-controllable device connected to the particular remote control.
  - H1 The TV is in the role of being turned on by the telecommand. It matters whether or not the TV is in the state ‘on’ or ‘off’; it changes this state upon receiving the telecommand.
  - H2 The TV is in a state such that it receives TV broadcast programs; in particular, it is a device that delivers the evening news. It is a device of mass media communication.
  - H3 The TV is in the role of the device that enables Amy to have her daily evening TV ritual. It serves Amy’s habits in a particular way.
  - H4 The TV and its program, when watched by Amy, makes it impossible to continue conversation with her. To Amy, the TV and its program is something that at this moment is more important than continuing her conversation.
  - H5 The TV and its program are a disruptive element to her friend’s interaction with Amy.

## 5.6 A Multitrack Notion of C-Constitution

I argued above that the cascade relations are second-order because they are relations between act-types, and therefore relations between, rather than within, first-order frames, in the frame-model adopted here. We now see that there is an even stronger argument for the second-order view: c-constitution between acts necessarily comes along with c-constitution of agencies and potential further arguments of the acts if they are shared across levels. These other tracks of c-constitution are conceptualized as roles of the arguments involved. Hence, c-constitution is a *multitrack* condition. Figure 6 displays a three-track sub-configuration cascade that would apply



**Fig. 6** Three tracks of c-constituency in a cascade

to the writing example. Notably, the parallel tracks in an action cascade intrinsically harmonize. To each of them the same circumstances—the “c” parameter of c-constitution—are relevant, and with them the level-specific contexts. The diagram highlights the multitude of c-const relations, the three tracks can alternatively be considered the components of one complex inter-level relation.

## 6 Reference and Composition

The assumption that action verb meanings are concepts with a cascade structure has far-reaching consequences not only for a theory of cognitive representation and decomposition, but also for the theory of reference and composition.

### 6.1 Meaning and Reference of the Verb Write

We call activities at all Levels H1 to H5 of the writing cascade “writing”, regardless if the higher levels are actually achieved. If we refer to a level higher than H1, a choice of alternative methods at Levels L and H1 is available, such as writing with a typewriter, or on a computer with a keyboard, on a smart phone with a touch screen etc. Thus, for present-day English, it is not to be assumed that the cascade in Fig. 4 represents the *lexical meaning* of the verb, as the lexical entry must not fix the method of writing. That does not mean that the level of the writing method is absent from the concept; it cannot be absent because it is required for logical reasons (there are

no higher-level acts without appropriate generating lower-level acts). Thus, I assume that the lexical meaning of the verb *write* is the cascade in Fig. 4 with the lowest level H1 and its generators left unspecified. In general, verbs for non-basic action *eo ipso* call for a lexical analysis in form of a cascade. If an unspecified generating level is addressed, for example by a modification of *write* with *shakily*, it is to be accommodated suitably.

The multilevel structure of the meaning is not a case of polysemy, that is, different senses on a par with each other. Rather, it is a case of *one* sense with several components, organized into a cascade. Of course, action verbs with a cascade structure meaning can be polysemous independently, requiring a separate cascade analysis for each sense.

When the verb *write* is used referentially, it refers to a whole cascade of act-TTs. Even if the very token of the verb is used in a way that relates to a specific level, for example, by specifying a product of a specific level or by applying level-specific modification, more than this level is concerned. On the one hand, reference is necessarily downward-complete: reference to a non-basic cascade level ontologically and conceptually requires generating act-TTs. This holds for all verbs that denote non-basic action: their cascade-format lexical meaning will contain at least one generating level, of an act-type which may or may not be specified. Even if unspecified, generating lower level actions are not of arbitrary type; rather they must be such that, under the circumstances one is entitled to assume, they level-generate what is at stake. On the other hand, we will further assume that, if a lower level is explicitly addressed, it will generate higher levels according to our assumptions about the circumstances. That does not mean we have to assume that always a complete writing cascade up to Level H5 is referred to. The circumstances may be such that they prevent level-generation of certain higher levels. Also, a given specification of the product argument, say as “whorls”, may preclude level-generation on the object track and therefore also on the act-track.

In addition to the levels subject to direct reference, we will be ready to generate further levels of a given TT cascade in our inevitable attempts to make more sense of what is said, by relating the act to further contexts in which it might matter. Thus, level-generation is a particularly rich source of conversational implicatures based on relevance. These cascade extensions will not be found in the lexical entries since they depend on the circumstances of an individual utterance.

## 6.2 Cascades and Composition

If we consider semantic meanings to be concepts, for example frame cascades for verbs of action, and if we are provided with explicit models of these concepts, we are in a position to ground a theory of semantic composition on decomposition. Semantic composition can then be modeled in more detail and more precisely. Also, if we know more about the meanings of words, we can start to model the interaction of semantic

information with context knowledge. Using the example of the verb *write*, I will illustrate some of the general perspectives of semantic composition emerging.

Let us assume we are to interpret a simple sentence with the verb *write* in finite use, with a subject and a direct object.

(35) *Martha wrote the statement.*

The lexical meaning of the name *Martha*, when taken as a person name, is a very simple frame: There is a central referential node typed as ‘person’ with one attribute, NAME, that carries the value ‘[Martha]’, basically an English sound and written form; we may add a GENDER attribute to the central node with the value ‘female’ if we consider it adequate to assume that bearer’s gender being female constitutes part of the meaning of the name *Martha*. The subject DP in (35) specifies the agent argument of the verb. Now, there are five agent nodes in the writing cascade that belong to an act typed as some level of writing. In principle, the frame for *Martha* can be unified with any one of them. What about the remaining four agent nodes? They will essentially be taken care of by the c-constitution requirements. In the simpler case of unsplit agency, Martha implements the agent at all levels, i.e. the scribbler, the scribe, the author, and the principal at the same time. If we allow for footing splits, the conditions are more involved: the level-agent is either Martha herself, or somebody who delegates this level to Martha or someone who Martha delegates this level to.

In addition to the full five-level readings of *write*, there is the possibility that the writing cascade may be implemented only up to a level lower than H5. Thus, there are three degrees of freedom given for the composition of verb and subject NP: (i) choice of the overall expansion of the writing cascade up to a level less than or equal H5; (ii) selection of a level for the agent; (iii) selection of the agent’s role in a footing structure. This amounts to a vast number of readings on this part alone.

Dealing with the direct object in (35) is less complex because the product is Level H5, an illocution. In order to be able to select the appropriate level for unifying the product node with the frame for *the statement*, we need to know that statements are illocutions, that is, we need an according frame representation of the noun *statement*. As to the remaining four object nodes in the cascade, again the c-const relation will take care; for any product at a Level  $n + 1$ , the product at Level  $n$  must support (i.e. c-constitute in the generalized sense) the higher-level product type. We may, however, also have product specifications that leave the type and level open, such as *it* or *that*. Depending on how the reference of the pronoun is determined in the given context, it might result in selecting a different level than was chosen for the agent. Therefore, the number of readings due to handling the agent argument potentially multiplies with the number of levels on account of level-selection for the object specification.

As is natural when one works with frames, I assume that the basic mechanism of semantic composition is unification.<sup>35</sup> Unification is restricted by the condition that

---

<sup>35</sup> According to the formal semantics view of composition, predicate expressions have open argument slots in their meaning to be “saturated” with the arguments. If we apply this view to the cascade approach, one level will be selected for the agent argument to be saturated and a possibly different

the type information on the nodes unified be compatible. In the case of level-specific object specifications or modifiers, this condition accounts for how these “find” their level to apply to. If there is more than one pair of nodes that fit, there may be more than one way of unification. We therefore have to accept that semantic composition is not deterministic. Although this is a bitter pill to swallow for some theoretical orientations in semantics, this consequence is after all welcome. All the readings possible are potentially “real”. If there are several readings to a construction, the compositional theory must predict all of them. Thus, the multilevel approach is on the one hand considerably more complex, but on the other able to account for the data more adequately.

The classical model of semantic composition is not a psychologically realistic model (and never was meant to be). In a realistic approach to semantic processing, the semantic agent will not only process linguistic information (i.e. syntactic structure and lexical meanings), but they will also draw on contextual knowledge *during* the process of composition, not only after it is finished (Hagoort et al. 2004). Aiming not at abstract sentence meaning, but at utterance meaning, i.e. meaning plus reference in the given context, the composing subject will merge the semantic information as early as possible with contextual information about the referents. For example, when faced with the sentence *Martha wrote the statement*, in a context where they know who Martha is, what statement is at issue, and which writing footing Martha can have, they may end up with one possible reading only. It is in this connection, where the dependence of c-constitution on the circumstances comes to bear crucially. The c-parameter in every cascade link *calls for* the inclusion of contextual knowledge in the compositional process; knowledge of the circumstances is necessary in order to decide which cascade levels are actually accomplished.

## 7 Conclusion: Cascades in Cognition, Semantics, and Life

We started out from Goldman’s (1970) theory of level-generation and act-trees. Taken as the psychological notion Goldman had in mind, level-generation provides the ground for a novel theory of the cognitive representation of action concepts: human action is conceptualized in multilevel cascade structures (the occasional basic acts notwithstanding). The levels of c-constitution are not levels of generality, but of constituency: lower-level acts constitute higher-level acts, where constituency is generally dependent on circumstances that make it possible.

In his introduction, Goldman relates his theory of action to the ontological debate about the question as to whether, say, flipping a switch and thereby turning on the light is one act or two. The problem dissolves, if one adopts the psychological view on the matter. From this perspective, Goldman’s theory is not about just act-tokens, but about act-tokens-of-a-type, i.e. what I dubbed “act-TTs”. There is no doubt

---

level for the product argument. The other agent slots and product slots are existentially saturated and imposed type conditions emanating from the c-const relations obtaining to the saturated nodes.

that, if one does something—one doing—one potentially enacts a whole cascade of action. All the acts in a cascade *really* are enacted; they *really* are as what they are categorized at each cascade level. This is reality *to us* as we cognitively construe the world. For psychology and for the analysis of verbal communication—and thereby for semantics and pragmatics—this is the relevant notion of reality.

In a second step, we applied Goldman's multilevel approach to action *verb* concepts in natural language. Almost all action verbs denote non-basic action and therefore cascades of action. Some examples of everyday activities such as writing or speech acts call for cascaded concepts of as much as six or more levels. Thus, the repertoire of natural language verb meanings provides ample evidence for a Goldmanian multilevel view on action categorization. As a theory of the structure of semantic verb concepts, the cascade approach has far-reaching consequences for semantic theory.

Linking the cascade theory of action to observations on the meanings of action verbs is not only an application of the theory; these observations conversely provide evidence for cognitive theory: if so many lexical verb concepts turn out to be multilevel, this must be due to the way in which our minds work.

A closer look at the participants in the acts within a cascade revealed that there are analogous constituency relationships between the respective participants at different levels. There is a track of stepwise upwards implementation of agency in terms of the finer-grained level-specific agent roles. A parallel track obtains for other participants involved through cascade levels. This finding suggests that the multilevel conceptualization of human action induces cascades not only for action itself, but also for agents and objects involved.

Can cascade theory be extended to other types of verb? One natural way of extension appears to be the generalization of c-constitution in a way that captures the meaning and relevance of arbitrary events for the options of acting. For example, a rainfall or a blackout or an insufficient battery stage of our mobile may *c*-constitute all sorts of conditions for possible and impossible action. The outcome of level-generation would be what events and situations *mean* to us and for our options to act. In any event, the findings on the multilevel categorization of action, as well as, derivatively, of roles to act in and roles in which objects may be involved in action suggest that the conceptualization of action may play a more fundamental and central role in our cognitive system than widely assumed.<sup>36</sup>

A radical induction from these findings might be this: All human categorization is, at least potentially, multilevel in the sense of cascade theory. Whatever we categorize, we categorize at potentially more than one level. This is owed to the fact that the bits and pieces of reality, or to be precise: of *what is reality to us as human cognitive subjects* always matter in many different contexts. The brief glimpse at upward cascading mechanisms in the verbal lexicon (Sect. 3.4) gave an impression of where cascading expands to: in many cases it is a projection into the realm of social action and interaction; in others, cascading takes categorization to the realm

---

<sup>36</sup>For a review of recent trends to the contrary in cognitive theory see Barsalou (2016): “Increasingly, researchers appreciate the central roles that action plays throughout cognition,” he concludes (p. 96).

of appraisal (with respect to personal or socially shared values). This might be taken as an indication that there be macrolevels across specific action types. Acquiring a vocabulary of verbs for human action with cascade structure meanings will help the members of a language community to synchronize their cascade level distinctions for single types of action as well as for overarching macrolevels. Clark's (1996) theory of language use is a detailed study of how conversational interactants synchronize their multilevel views of the interaction they are engaged in.

The higher levels of an action cascade can be considered as corresponding to as many respects in which the doing has *meaning* to us (in a nontechnical sense). Likewise, persons in roles matter at the level of action that defines this role, and so do objects involved in action. Conversely, acts, persons, and objects can be viewed as lacking meaning to us as long as they, for us, do not c-constitute anything at a higher level. Of course, what carries meaning to a subject is first of all a personal issue. There are, however, socially established ways of c-constitution that will be anticipated by persons in social interaction (cf., for example, Searle's (1995) social ontology).

An aspect of cascade theory that was not discussed here is the role of cascades in practical knowledge. The basic levels of cascades, like pressing a button on a remote control, flipping a switch, touching a symbol on a touch screen, constitute the methods we learn and then command for doing the higher-level types of action such as turning on the TV, or the light, or starting an app. In our complex and ever-expanding knowledge-how about the world we live in, we have learned countless such cascades from our earliest stages of life on: we have learnt by which methods to do what. Notably, most of the time, we have no understanding of the underlying circumstances and causal relations responsible for the possibility of these level-generations; for all practical purposes, they are just given in our world and part of it. Level-generation in these cases does not seem to involve any kind of reasoning. Thus, the observation that most of our practical knowledge about the environment has cascade structure constitutes solid evidence that level-generation, or c-constitution, is indeed a fundamental brain mechanism, as I assumed above. This view of the role of cascade formation in the psychology of knowing how and learning by doing is developed in the contribution by Kalenscher et al. in this volume. That contribution is about rats, suggesting that cascade theory might apply even to animal cognition.

**Acknowledgements** The research for this chapter was supported by the Collaborative Research Center 991 "The Structure of Representations in Language, Cognition, and Science" financed by the German Research Foundation (Deutsche Forschungsgemeinschaft) and Heinrich-Heine-Universität Düsseldorf. I am deeply indebted to Henk Zeevat who drew my attention to Goldman's and Clark's work. I owe very much to discussions on cascade theory to Henk Zeevat and my collaborators Curt Anderson, Ekaterina Gabrovska, and Wilhelm Geuder; Liu Fan, Nanjing University, China, helped me with the Mandarin data. I profited a lot from detailed comments by Lawrence Barsalou and four anonymous reviewers.

## References

- Andreou, M., & Petitjean, S. (this volume). An XMG account of multiplicity of meaning in derivation.
- Anscombe, G. (1963). *Intention* (2nd ed.). Oxford: Blackwell.
- Asher, N. (2011). *Lexical meaning in context: A web of words*. Cambridge: Cambridge University Press.
- Austin, J. L. (1962). *How to do things with words*. Cambridge, MA: MIT Press.
- Balogh, K., & Osswald, R. (this volume). A frame-based analysis of verbal particles in Hungarian.
- Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In A. Lehrer & E. F. Kittay (Eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization* (pp. 21–74). Hillsdale, NJ: Lawrence Erlbaum Associates.
- 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. F. Friston, & D. Kragic (Eds.), *The pragmatic turn: Toward action-oriented views in cognitive science* (pp. 81–98). Cambridge, MA: MIT Press.
- Bennett, J. (1988). *Events and their names*. Oxford: Oxford University Press.
- Bücking, S. (2014). Elaborating on events by means of English *by* and German *indem*. In C. Piñón (Ed.), *Empirical issues in syntax and semantics* (Vol. 10). <http://www.cssp.cnrs.fr/eiss10/>.
- Carpenter, B. (1992). *The logic of typed feature structures*. Cambridge: Cambridge University Press.
- Castañeda, H.-N. (1979). Intensionality and identity in human action and philosophical method. *Noûs*, 13, 235–260.
- Chang, S.-W. (Ed.). (2016). *The Routledge encyclopedia of the Chinese language*. London, New York: Routledge.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Davidson, D. (1963). Actions, reasons, and causes. *The Journal of Philosophy*, 23, 685–700.
- Dowty, D. R. (1979). *Word meaning and Montague Grammar*. Dordrecht: Reidel.
- Engelberg, S. (2005). Stativity, supervenience, and sentential subjects. In C. Maienborn & A. Wöllstein (Eds.), *Event arguments: Foundations and applications* (pp. 45–68). Tübingen: Niemeyer.
- Fleischhauer, J. (2016). *Degree gradation of verbs*. Düsseldorf: Düsseldorf University Press. <http://dup.oa.hhu.de/522/>.
- Gamerschlag, T., Geuder, W., & Petersen, W. (2014). Glück auf, der Steiger kommt: A frame account of extensional and intensional *steigen*. In D. Gerland, C. Horn, A. Latrouite, & A. Ortmann (Eds.), *Meaning and grammar of nouns and verbs, Studies in language and cognition* (Vol. 1, pp. 115–144). Düsseldorf: Düsseldorf University Press. Open access available at <http://dup.oa.hhu.de/517/>.
- Gamerschlag, T., & Petersen, W. (this volume). On the fictive reading of German *steigen* ‘climb, rise’: A frame account.
- Ginet, C. (1990). *On action*. Cambridge: Cambridge University Press.
- Goffman, E. (1979). Footing. *Semiotica*, 25, 1–29. Also published in Goffman, E. (1981). *Forms of talk*, pp. 124–157. Philadelphia, PA: University of Pennsylvania Press.
- Goldman, A. I. (1970). *Theory of human action*. Princeton, NJ: Princeton University Press.
- Goldman, A. I. (1979). Action, causation, and unity. *Noûs*, 13, 261–270.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438–441.
- Kalenscher, T., Schönfeld, L.-M., Löbner, S., Wöhr, M., van Berkel, M., Zech, M.-P., van Wingerden, M. (this volume). *Rat ultrasonic vocalizations as social reinforcers—Implications for a multilevel model of the cognitive representation of action and rats' social world*.
- Kallmeyer, L., & Osswald, R. (2013). Syntax-driven semantic frame composition in lexicalized tree adjoining grammars. *Journal of Language Modelling*, 1, 267–330.
- Kearns, K. (2003). Durative achievements and individual-level predicates on events. *Linguistics and Philosophy*, 26, 595–635.

- Levin, B. (1993). *English verb classes and alternations*. Chicago: The University of Chicago Press.
- Levinson, S. C. (1988). Putting linguistics on a proper footing: Exploration in Goffman's concepts of participation. In P. Drew & A. Wootton (Eds.), *Erving Goffman: Exploring the interaction order* (pp. 161–227). Cambridge: Polity Press.
- Löbner, S. (2013). *Understanding semantics* (2nd ed.). London, New York: Routledge.
- Löbner, S. (2014). Evidence for frames from natural language. In T. Gamerschlag, D. Gerland, R. Osswald, & W. Petersen (Eds.), *Frames and concept types* (pp. 23–68). Heidelberg, New York: Springer.
- Löbner, S. (2017). Frame theory with first-order comparators: modeling the lexical meaning of punctual verbs of change with frames. In H. H. Hansen, S. E. Murray, M. Sadrzadeh, & H. Zeevat (Eds.), *Logic, language, and computation. Proceedings of the 11th international Tbilisi symposium* (pp. 98–117). Heidelberg, New York: Springer.
- Martin, S. E. (1975). *A reference grammar of Japanese*. New Haven, London: Yale University Press.
- McCann, H. J. (1982). The trouble with level generation. *Mind*, 91, 481–500.
- McCawley, J. D. (1968). Lexical insertion in a transformational grammar without deep structure. *CLS*, 4, 71–80.
- McLaughlin, B., & Bennett, K. (2014). Supervenience. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2014 Ed.). Retrieved January 10, 2018, from <https://plato.stanford.edu/archives/spr2014/entries/supervenience/>.
- Moltmann, F. (2017). Levels of linguistic acts and the semantics of saying and quoting. In S. L. Tsohatzidis (Ed.), *Interpreting Austin: Critical essays* (pp. 34–59). Cambridge: Cambridge University Press.
- Naumann, R. (2013). An outline of a dynamic theory of frames. In G. Bezhanishvili, S. Löbner, V. Marra, F. Richter (Eds.), *Proceedings of the 9th international Tbilisi symposium* (pp. 115–137). Heidelberg: Springer.
- Petersen, W. (2007). Decomposing concepts with frames. In J. Skilters, F. Toccafondi, G. Stemberger (Eds.), *Complex cognition and qualitative science. The Baltic international yearbook of cognition, logic and communication* (Vol. 2, pp. 151–170). Riga: University of Latvia.
- Pustejovsky, J. (2009). The semantics of lexical underspecification. *Folia Linguistica*, 32(3–4), 323–348. <https://doi.org/10.1515/flin.1998.32.3-4.323>.
- Sæbø, K. J. (2008). The structure of criterion predicates. In J. Dölling, T. Heyde-Zybatow, & M. Schäfer (Eds.), *Event structures in linguistic form and interpretation* (pp. 127–147). Berlin: Mouton de Gruyter.
- Sæbø, K. J. (2016). "How"-questions and the manner–method distinction. *Synthese*, 193, 3169–3194.
- Searle, J. R. (1995). *The construction of social reality*. New York: The Free Press.
- Stiebels, B. (1996). *Lexikalische Argumente und Adjunkte: zum semantischen Beitrag von verbalen Präfixen und Artikeln*. Berlin: Akademie Verlag.
- Tsai, W.-T. D. (2012). *High applicatives are not high enough: A cartographic solution*. Paper presented in the 14th China International Conference on Contemporary Linguistics, Xi'an Foreign Language University, Xi'an, May 2012. Retrieved January 28, 2018, from <http://ling.auf.net/lin/gbuzz/001062/current.pdf>.
- Van Valin, R. D., & LaPolla, R. J. (1997). *Syntax. Structure, meaning and function*. Cambridge: Cambridge University Press.
- Wegener, H. (1985). *Der Dativ im heutigen Deutsch*. Tübingen: Gunter Narr Verlag.
- Zinova, J. (2016). *Russian verbal prefixes: A frame semantic analysis*. Ph.D. dissertation. Düsseldorf: Heinrich Heine Universität.

## ***Online Corpora***

BCC Beijing Language and Culture University Chinese Corpus. <http://bcc.blcu.edu.cn/lang/en>.

BNC BNCweb. British National Corpus online. <http://bncweb.lancs.ac.uk/>.

DWDS Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart. <https://www.dwds.de/>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# **Prototypes and Probabilities**

# Modification and Default Inheritance



Corina Strößner, Annika Schuster, and Gerhard Schurz

**Abstract** Modification usually decreases the judged likelihood of typicality statements. People judge “Old coyotes howl” as less likely than just “Coyotes howl”. This paper addresses this so-called modification effect. In order to analyse the effect, we propose an extended modification model based on the selective modification model by Smith et al. (1988) and Barsalou’s (1992) frames. In this model we introduce cross-attributional constraints that explain how a change in one dimension leads to an alteration of another attribute, especially if the modifier is not typical. Finally, we discuss data from Connolly et al. (2007) and present new experimental evidence from an explorative study.

**Keywords** Modifier effect · Constraints · Frames · Prototype theory · Compositionality

## 1 Prototype Compositionality and Modification

Originating in the work of Eleanor Rosch and her co-authors (Rosch and Mervis 1975; Rosch et al. 1976; Rosch 1978), the prototype theory of concepts influenced the way psychologists, linguists and philosophers understand concepts enormously. In its most popular version, prototype theory claims that concepts are associated with internal typicality orderings. This thesis is well-confirmed and widely accepted. It is also well-known that human agents are capable of composing concepts. But how can prototype theory contribute to the understanding of this creative process? Typicality doesn’t combine in a straightforward way. A typical pet fish is neither a typical fish nor a typical pet (cf. Fodor and Lepore 1996). Elaborated models of prototype composition have been developed since the 1980s. Hampton (1987) discusses how

---

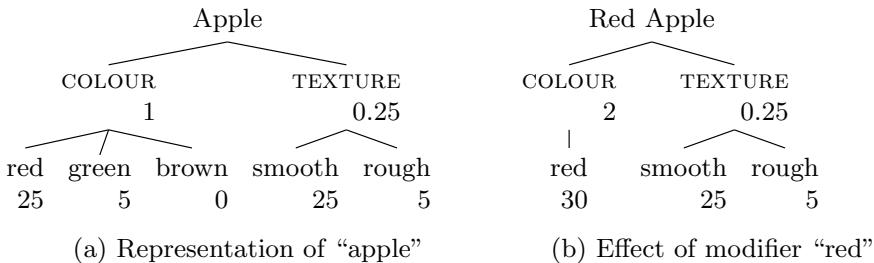
C. Strößner (✉)

Ruhr University Bochum, Bochum, Germany

e-mail: [Corina.Stroessner@ruhr-uni-bochum.de](mailto:Corina.Stroessner@ruhr-uni-bochum.de)

A. Schuster · G. Schurz

Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany



**Fig. 1** Modification in SMM, following (Smith et al. 1988, pp. 490, 494)

the typicality ratings of noun constructions like “sports that are also games” are determined by the importance of the properties for their components. The selective modification model proposed by Smith et al. (1988), on the other hand, concerns modifications that are realized in adjective-noun combinations. These are at the focus of this paper.

The selective modification model, henceforth SMM, starts with a representation of prototype concepts as attribute-value structures, an importance measure for attributes, called diagnosticity,<sup>1</sup> and a voting for the values, called salience (cf. Smith et al. 1988, p. 489).

Modification is understood as a strictly selective process in the SMM, the effect of which is limited to one attribute. The modifier selects the attribute the adjective addresses, shifts all votes to its value, and increases the importance of this particular attribute (cf. Smith et al. 1988, p. 492). Figure 1 shows how the modifier “red” for “apple” operates on the colour attribute: all votes go to “red” and the importance of colour is increased. The SMM is a very simple but still effective approach to prototype compositionality. However, it also has several limitations. The most important one is the strictness of its selectivity. This strong assumption prevents modification from altering anything but one attribute. Thus, SMM predicts that all non-modified properties are inherited. Smith et al. (1988, p. 497) are aware of possible correlations between attributes, but they defer necessary adjustments to a subsequent cognitive process.

Connolly et al. (2007) presented experimental evidence that is not compatible with the predictions of SMM: subjects rate unmodified statements like “Ravens are black” more likely than modified ones like “Feathered ravens are black”. The judged likelihood is even lower if the modifiers are not typical, e.g., for “Jungle ravens are black”, and further decreases if two modifiers are used, as in “Young jungle ravens are black”. On a scale from 1 (very unlikely) to 10 (very likely) the mean rating was 8.36 for unmodified sentences (A), 7.71 for typical modifications (B), 6.91 for non-typical ones (C) and only 6.48 for double modifications (D) (cf. Connolly et al. 2007, p. 11f). Jönsson and Hampton (2012) and Hampton et al. (2011) confirmed

<sup>1</sup>The term “diagnosticity” is also often used to indicate the specificity of a property. Thus we prefer the expression “attribute importance”.

these findings in further experiments. Gagné and Spalding (2011, 2014) observed a similar effect for meaningless pronounceable modifiers. While the existence of the effect is uncontroversial, there is a lively debate on its interpretation.

Connolly et al. (2007) claim that their experiment proves that people don't use a default to prototype strategy. Typical values are not inherited by subcategories but rather inferred in post-compositional step, which is largely lead by personal knowledge (cf. Connolly et al. 2007, p. 15). Jönsson and Hampton (2012, p. 109), on the contrary, argue that typical properties are inherited. Only in a second step, subjects decrease their certainty about the typical properties, mostly because of background knowledge or for pragmatic reasons. Gagné and Spalding (2014) take a third position. According to them, typical properties are not inherited but inferred from the meta-knowledge that subcategories resemble the category to some degree but are still distinct (cf. Gagné and Spalding 2014, p. 1291).

The authors agree in taking their results to be incompatible with the SMM. To begin with, it is hard to explain why modifications influence typical values of attributes that are not addressed by the modification. On top of that, the remarkable difference between typical modifiers and other modifiers is an unexplainable mystery to the SMM (cf. Jönsson and Hampton 2012, p. 111). However, the SMM also explains some results. The rating of a modified sentence is highly correlated with the typicality rating of its unmodified counterpart (cf. Jönsson and Hampton 2012, p. 98). The contribution of the head noun to the modification occurs even if the modifier is not a meaningful word, and thus certainly not learned from experience (cf. Gagné and Spalding 2014, pp. 1287–1288). In sum, experimental evidence has revealed three stable effects for a head noun S, its prototypical property P and the modifier M:

1. “S is P” is usually rated as more likely than “M S is P”
2. There is a positive correlation between the rated likelihood of “S is P” and “M S is P”.
3. The loss in rated likelihood from “S is P” to “M S is P” is smaller if M is typical for S.

The modification effect doesn't depend on how central the typical property P is. Hampton et al. (2011) even produced the effect for properties that are analytically true, like being a bird for raven.

While a reference to post-compositional adjustments can save the basic idea of SMM, it also reduces its empirical content and strength. This is why we propose to enrich the SMM by making use of frames (Barsalou 1992), i.e. recursive attribute-value structures that allow the specification of constraints between values of different attributes. This is carried out in the next section. In the third section, we show that an application of our model to data from Connolly et al. (2007) shows a stronger decrease in likelihood in the presence of constraints. We finally present new experimental evidence we gathered in an exploratory study.

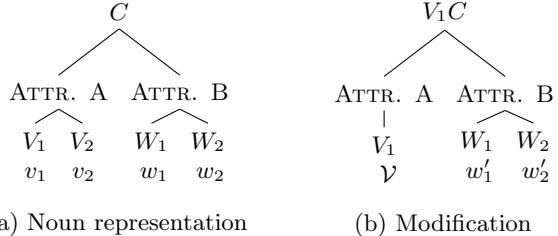
## 2 An Extended Modification Model

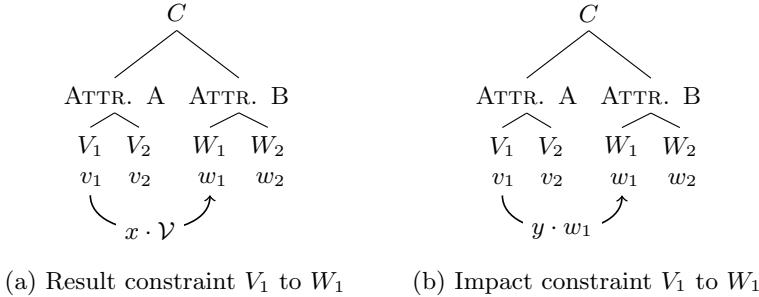
We understand modification as an asymmetrical composition that is usually realized in adjective-noun compounds. Depending on the way the modifier interacts with the head noun, normal modifications can be distinguished from deviant forms of modification. In a normal modification, the modifier picks a value for a attribute in the noun frame. Deviant and privative modifications like “stone lion”, on the other hand, are grammatically like normal modifications but interfere with the noun in a more drastic way. They often lead to coercion, metaphorical use or to a high demand for context. When we are confronted with such compounds we have to reconsider our normal interpretation of the head noun. Understanding of deviant modifications confronts us with its own obstacles, the reasons for which do not lie in prototype theory. Our approach is therefore focused on the understanding of normal modifications.

For our illustration of modification we refer to the frame model of Barsalou (1992), who claims that conceptual content is best represented in terms of attribute-value structures. Cross-attributional dependencies are illustrated as constraints, i.e. as relations between values. A constraint can for example state that a green colour of an apple indicates sour taste. Barsalou’s frames comprise the attribute-value structure of SMM and, additionally, they allow the representation of dependencies between values of different attributes by means of constraints.

Our enriched model of modification states, like SMM, that a modifier specifies a value of a noun’s attribute and shifts all votes to this value. SMM also claims an importance boost of the according attribute. Although we readily accept this thesis, it will not matter in the following discussion. Our focus is on the changed likelihood of values, i.e. the shift of votes. Thus, we ignore importance measures in this paper. Our essential extension of the SMM is the *constraint thesis*, which contradicts the strict selectivity of SMM: by modification, the selected value collects all votes *and activates constraints to other values*. The constraint thesis will be formalised in the next section. The discussion is based on the minimal model, shown in Fig. 2. We consider a concept  $C$  with two attributes, A and B. The values of A are  $V_1$  and  $V_2$  with the respective votes  $v_1$  and  $v_2$ . The attribute B has the values  $W_1$  with  $w_1$  votes and  $W_2$  with  $w_2$  votes (Fig. 2a).  $\mathcal{V} = v_1 + v_2 = w_1 + w_2$  is the number of total votes.

**Fig. 2** Basic model



**Fig. 3** Constraints

$V_1$  is the value of the modifier, as shown in Fig. 2b , with the new votes  $v'_1 = \mathcal{V}$  for  $V_1$ ,  $v'_2 = 0$  for  $V_2$  and the new votes  $w'_1$  for  $W_1$  as well as  $w'_2$  for  $W_2$ .

### Typicality of values and modifiers

The typicality of a modifier has a well-documented influence in all experiments. The existing literature, starting from Connolly et al. (2007), distinguishes between modifiers that are typical and other modifiers.<sup>2</sup> We will refine the notion of typicality. Drawing on the distributions of votes on an attribute, we distinguish *typical* values with a very high proportion of votes from *atypical* values with a very low proportion of votes. Values with a medium number of votes are called *neutral*.

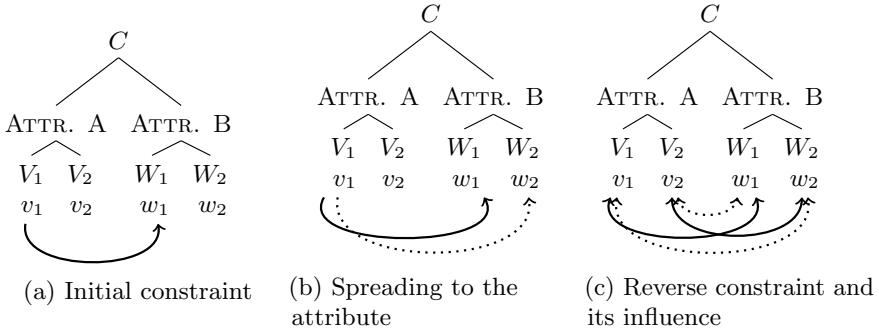
### Bringing constraints to SMM

Since the SMM is based on quantitative specifications, especially votes for values, it is necessary to quantify constraints as well. There are different ways to achieve this.

One possibility to quantify constraints is to specify the proportion of votes that will be given to the target value if the constraining value comes to be known. An example of such a result constraint is given in Fig. 3a:  $x$  with  $0 \leq x \leq 1$  is the proportion of votes the result constraint from  $V_1$  gives to  $W_1$ . The value  $x$  can be interpreted as the conditional probability of  $W_1$  given  $V_1$ . This allows us to tie on to results in the field of probability theory.<sup>3</sup> If no further constraint is involved, the other votes on B ( $w_2, w_3, \dots, w_n$ ) need to be adjusted in a way that reflects their initial proportion, namely as  $w'_i = \frac{w_i}{\mathcal{V}} \cdot (\mathcal{V} - w'_1)$ . The strength or impact of the constraint is apparent by the difference between the initial votes and the new votes.

<sup>2</sup>They refer to typical modifiers as those properties that were collected in feature lists by Cree and McRae (2003).

<sup>3</sup>An elaborated way to model dependencies probabilistically is the theory of Bayesian nets introduced by Pearl (1998). For our basic model, which is based on correlations, Bayesian nets are overly powerful.



**Fig. 4** Influence of constraints

An alternative approach are impact constraints that specify the alteration of the votes by the particular constraint. In the impact representation of a constraint from  $V_1$  to  $W_1$ , we give a factor  $y$  such that  $0 \leq y \leq \frac{1}{w_1/\mathcal{V}}$ , which is multiplied with  $w_1$  as illustrated in Fig. 3b. The new votes for  $w_2, w_3, \dots, w_n$  after activating the constraint from  $V_1$  to  $W_1$  are calculated as  $w'_i = w_i \cdot \frac{\mathcal{V} - w'_1}{\mathcal{V} - w_1}$ . The direction of the constraint is now apparent in the constraint itself: For positive constraints we have  $y > 1$ , while for negative constraints  $y < 1$ . Neutral constraints with  $y = 1$  can be used to represent known irrelevance.

The influence of constraints spreads. Any active constraint from  $V_1$  to  $W_1$  has an influence on  $W_1$ 's alternatives. If  $V_1$  increases the likelihood of  $W_1$ , then it decreases the likelihood of its alternatives, e.g.,  $W_2$ , and the other way around. Furthermore, the constraint from  $V_1$  to  $W_1$  leads to a constraint from  $W_1$  to  $V_1$ . It can be calculated by Bayes' theorem as  $P(V_1|W_1) = \frac{x \cdot (v_1/\mathcal{V})}{w_1/\mathcal{V}}$ . Thus,  $W_1$  increases the likelihood of  $V_1$  and decreases the likelihood of  $V_2$  if and only if  $V_1$  increases the likelihood of  $W_1$ . This is illustrated in Fig. 4b, where solid arrows indicate increased likelihoods (positive constraints) and dotted ones indicate decreased likelihoods (negative constraints).

## Constraining constraints

Constraints are restricted. For example, a typical value cannot severely increase the likelihood of an atypical one. In order to determine the impact of a constraint, we introduce the factor  $f$  that is needed to shift all votes to the constraining value  $V_1$ , i.e. to make it maximally probable:  $f = \frac{1}{P(V_1)}$ . To approach the possible influence on  $W_1$ , we rely on  $P(W_1)' = f \cdot P(W_1 \wedge V_1) + 0 \cdot P(W_1 \wedge \neg V_1)$ . For a positive constraint, we stipulate that  $P(W_1 \wedge V_1)$  is as high as possible. Thus,  $f = \frac{1}{P(V_1)}$  is also the maximal positive impact a constraint from  $V_1$  can have, of course still with the limit that  $P(V_1) \cdot f \leq 1$ . For calculating the maximal negative constraint we assume  $W_1 \wedge \neg V_1$  to be as likely as possible.  $P(W_1 \wedge \neg V_1)$  cannot be larger than  $P(\neg V_1) = 1 - P(V_1)$ , i.e.  $P(W_1 \wedge V_1) \geq P(W_1) - P(V_1)$ .

**Table 1** Possible constraints and their results

Typicality of values	$v_1/\mathcal{V}$	$w_1/\mathcal{V}$	$P(V_1)$	$P(W_1)$	$P(W_1 V_1)$		Maximal votes change	
					Max	Min	Gain	Loss
Atypical to atypical	3/30	3/30	$\frac{1}{10}$	$\frac{1}{10}$	1	0	+27	-3
Atypical to neutral	3/30	15/30	$\frac{1}{10}$	$\frac{1}{2}$	1	0	+15	-15
Atypical to typical	3/30	27/30	$\frac{1}{10}$	$\frac{9}{10}$	1	0	+3	-27
Neutral to atypical	15/30	3/30	$\frac{1}{2}$	$\frac{1}{10}$	$\frac{1}{5}$	0	+3	-3
Neutral to neutral	15/30	15/30	$\frac{1}{2}$	$\frac{1}{2}$	1	0	+15	-15
Neutral to typical	15/30	27/30	$\frac{1}{2}$	$\frac{9}{10}$	1	$\frac{4}{5}$	+3	-3
Typical to atypical	27/30	3/30	$\frac{9}{10}$	$\frac{1}{10}$	$\frac{1}{9}$	0	+0.33	-3
Typical to neutral	27/30	15/30	$\frac{9}{10}$	$\frac{1}{2}$	$\frac{5}{9}$	$\frac{4}{9}$	+1.66	-1.66
Typical to typical	27/30	27/30	$\frac{9}{10}$	$\frac{9}{10}$	1	$\frac{8}{9}$	+3	-0.33

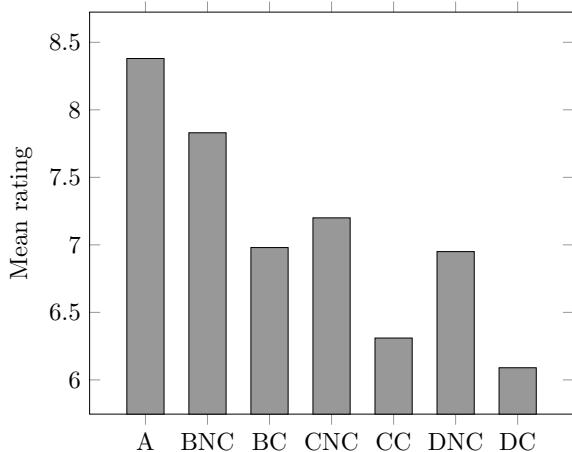
Table 1 shows how the rules restrict the effect of constraints. The initial votes on the modifying value  $V_1$  play a crucial role. If  $V_1$  is rather atypical, i.e. in the first three combinations, then the constraint can change the new distribution of votes severely. A typical modifier  $V_1$ , on the other hand, has only a limited potential to alter the initial distribution of votes.

Besides the formal considerations, there are conceptual restrictions. Prototype concepts represent property clusters (Rosch and Mervis 1975; Schurz 2012). Within the supercategory, typical values of a prototype concepts are positively correlated with each other. This correlation is not always inherited by the subcategories: Within the class of vertebrates, a beak is a good predictor of flying-ability but not in the category of birds. However, the positive correlation often remains valid, if functionality is involved. The beat of the heart is causally related with almost all vital properties of an organism and thus also statistically correlated with them. The typical shape of a tool is adjusted to its typical purposes. Positive associations between typical values in a category are frequent. For the formal reasons explicated above, these constraints should not be expected to lead to a high variability in modifications. However, their negative counterparts for atypical values are quite effective. Applying “biped” to “human” has little effect on expectations about moving abilities, while applying “non-biped” has a crucial influence.

### 3 Experimental Data

The introduced extended modification model predicts that the occurrence and the direction of alteration by a modifier is determined by the existence of positive and negative constraints and that less typical modifiers result in larger changes. We contend that this is the rational way to handle the information one has about noun and

**Fig. 5** Mean ratings without constraint (NC) and with constraint (C) by modifier condition



modifier. We investigated whether people follow this strategy by a further analysis of the data from Connolly et al. (2007) and in an exploratory study we carried out.

### 3.1 Constraint Influences in the Data of Connolly et al. (2007)

If our extended modification model is accurate, it should be possible to find influences of constraints on the likelihood of modified sentences in the original data set by Connolly et al. (2007).<sup>4</sup> Our research group thus examined the original stimuli and agreed on constraints between modifiers and ascribed properties. A similar idea can be found in Jönsson and Hampton (2012), where the *subjects* were asked to justify higher or lower likelihood ratings of modified sentences. The main reasons given were pragmatic (e.g., the weirdness of the modified sentences), justifications by background knowledge about the modifier, or uncertainty about the modified noun.

We determined constraints for 5 B-modifiers, 14 C-modifiers and 21 D-modifiers. The mean ratings for constrained and unconstrained sentences by question type are shown in Fig. 5. The decrease in judged likelihood is much stronger for the constrained sentences. However, this result has to be interpreted keeping in mind that our post hoc analysis results in different sample sizes (e. g. 350 ratings for the unconstrained B-condition compared to 50 ratings for the B-constraint condition).

Since modifications do not necessarily decrease, but in some cases may increase the likelihood of a property (compare “Hamsters live in cages” and “Pet hamsters live in cages.”), it makes sense to look at the absolute values of the differences to the baseline conditions A, shown in Table 2. Here, the difference between constrained

<sup>4</sup>This analysis was made possible because the authors kindly provided us with their original data.

**Table 2** Mean absolute differences from the baseline condition A without and with constraint and in total

	A-B	A-C	A-D
No constraint	0.78	1.123	1.405
Constraint	1.46	2.3	2.319
Total	0.856	1.535	1.885

**Table 3** Results of post-hoc significance test (insignificant results shaded)

	A	B-NC	C-NC	D-NC	B-C	C-C
B-NC	0.010					
C-NC	0.000	0.007				
D-NC	0.000	0.000	0.996			
B-C	0.000	0.175	1.000	1.000		
C-C	0.000	0.000	0.002	0.151	0.728	
D-C	0.000	0.000	0.000	0.001	0.160	1.000

and unconstrained versions is even more obvious: the reduced likelihood is nearly twice as high for the constrained sentences. Furthermore, there is almost no difference between the simple (C) and double (D) modification, indicating that constraints have a stronger influence on the judged likelihood than modification.

The results of t-tests between all groups with Hochberg's GT2 correction (for different sample sizes) are shown in Table 3. All groups differ significantly from the baseline condition A. The differences between constrained and unconstrained sentences are significant ( $p < 0.01$ ), except for the constrained B-condition, which is likely explained by its small sample size. The differences between the C- and D-conditions are not significant, neither for the constrained nor for the unconstrained version.<sup>5</sup> The results indicate that a more accurate grouping of the sentences would be between constrained and unconstrained modifications and neglecting the effect of double modification.

These analyses only allow for tentative conclusions because of the different sample sizes. But we can see a clear tendency in accordance with the predictions of the extended modification model: the change in likelihood ratings in the original data was shown to be much more distinguished for sentences in which the chosen modifiers constrain the assigned property.

<sup>5</sup> Jönsson and Hampton (2012, p. 98) also found insignificant differences between C- and D-stimuli in post-hoc pairwise comparisons.

### 3.2 Experiments

In order to test several of our empirical predictions we designed an *exploratory study* with few items and a comparatively small group of subjects. The described experiment served as a preparation of a larger study, reported elsewhere (Strößner and Schurz 2020). We tested several question types on four items.

#### 3.2.1 Method

##### Participants

Subjects were 48 students of the Heinrich-Heine-Universität Düsseldorf, who were paid for participation.

##### Material

We used German translations of four items from Connolly et al. (2007). Two of them were previously judged to have no constraint between modifier and ascribed property by the members of our research group. For the third and fourth item, the modifiers were suspected to have a constraint on the typical property. This pre-experimental classification by the authors was used in order to look whether items with a suspected knowledge constraint behave differently. Previous studies by (Jönsson and Hampton 2012) have shown that subjects are often aware of subtle dependencies between modifier and property if they have to justify a lower likelihood rating for the modified sentence. It has been noted that these justifications could also be made up only after the rating task rather than really influencing it (cf. Gagné and Spalding 2014, p. 1290). Moreover, we were also interested to know whether knowledge constraints are purely subjective or intersubjective. If constraints are purely subjective, then there should be little differences between the items with constraint and the items without constraint. In addition to the preclassification by the authors, we also gathered relevance ratings from the subjects, which will be reported below. The double-modification was only tested for the two items with presumed non-relevant modifiers. The items are shown in list 1. The questions types are listed in list 2. Subjects gave ratings on the typicality and likelihood of the items (question type P and T) as well as on the typicality and likelihood of the modifier (question type PM and TM). The relevance rating was gathered with question type RM.

### 1. Lambs

- A Lämmer sind weiß. (Lambs are white.)
- B Flauschige Lämmer sind weiß. (Fluffy lambs are white.)
- C Norwegische Lämmer sind weiß. (Norwegian lambs are white.)
- D Langhaarige, norwegische Lämmer sind weiß. (Long-haired Norwegian lambs are white.)

### 2. Shirts

- A Hemden haben Knöpfe. (Shirts have buttons.)
- B Baumwollhemden haben Knöpfe. (Cotton shirts have buttons.)
- C Kratzige Hemden haben Knöpfe. (Itchy shirts have buttons.)
- D Kratzige Leinenhemden haben Knöpfe. (Itchy canvas shirts have buttons.)

### 3. Limousines

- A Limousinen sind lang. (Limousines are long.)
- B Teure Limousinen sind lang. (Expensive limousines are long.)
- C Preisgünstige Limousinen sind lang. (Inexpensive limousines are long.)

### 4. Sofas

- A Sofas stehen im Wohnzimmer. (Sofas are in living rooms.)
- B Bequeme Sofas stehen im Wohnzimmer. (Comfortable sofas are in living rooms.)
- C Unbequeme Sofas stehen im Wohnzimmer. (Uncomfortable sofas are in living rooms.)

#### **List 1** Items used in our experiment

- P Subjects rated the likelihood of the property for the unmodified and modified nouns.
- T Subjects rated the typicality of the property of the unmodified and modified nouns.
- PM Subjects rated the likelihood of the modifiers for the nouns.
- TM Subjects rated the typicality of the modifiers for the nouns.
- RM Subjects rated whether the modified attribute is relevant for the target attribute.

#### **List 2** Question types

### Design

In the first questionnaire, subjects were instructed to answer how typical they rate the default property, e.g. *being long*, for the modified and unmodified nouns, e.g. *limousines*, *expensive limousines* and *inexpensive limousines* (question type T). One group of 19 participants rated the items 1 and 3. Another group of 19 subjects rated items 2 and 4. Both groups also rated the typicality of the modifiers, e.g. *being*

*expensive* and *being inexpensive* for limousines (question type TM). In the second questionnaire, the subjects of both groups were asked to rate the likelihood of the same items (question type P and PM). The likelihood ratings and typicality ratings were gathered in two separate questionnaires but came from the same subjects. The unmodified and modified conditions as well as the rating of the modifiers were mingled but appeared on the same questionnaire. The participants thus saw their own answers and were potentially able to review and revise them. The last questionnaire contained relevance ratings (question type RM) for all items and modifiers. Subjects rated whether the modified attribute is relevant for the target attribute, e.g. whether the length of a limousine is related to its price. All judgements were given on a scale from 0 to 10. For the relevance question, subjects had the possibility to answer “I don't know”.

### 3.2.2 Results

#### Typicality and Probability

In our model, probability plays a crucial role for defining typicality. We thus tested, whether the typicality ratings and the likelihood ratings are similar. This question is also important because even in a probabilistic approach, there are different notions of typicality: Schurz (2012) distinguishes *typicality in the wide sense* as probability in the category from *typicality in the narrow sense*, where a property has also to be improbable in sibling categories, i.e. highly discriminatory. This second criterion is what Rosch (1978) terms *cue validity*. For example, having a heart is only typical in the wide sense for birds. Having a beak is also typical in the narrow sense. Typicality in the wide sense justifies prediction of properties from known membership. Typicality in the narrow sense also allows to infer membership from known properties.

Table 4 shows the frequencies of the difference of all typicality ratings compared to the respective likelihood ratings. The typicality and likelihood ratings were very similar. In more than half of the pairs, they were even rated exactly the same.

**Table 4** Likelihood compared to typicality: cases and percentages

	+8	+7	+6	+5	+4	+3	+2	+1	0	-1	-2	-3	-4	-5
2	2	1	3	8	14	24	37	64	355	77	33	10	5	10
%	0.3	0.2	0.5	1.2	2.2	3.7	5.8	10	55.2	12	5.1	1.6	0.8	1.6

Paired sample t-tests for the 24 pairs showed that only four pairs were significantly different.<sup>6</sup> The result strongly indicates that subjects preferred a wider notion of “typical” in the task. This supports our definitions of typicality in terms of probability.

### The status of the modifier

The data on the typicality and likelihood of the modifier in relation to the head noun allowed us to confirm that the B condition modifiers were also considered as typical by subjects in the German speaking community in comparison to the modifiers in the C and D condition. The mean values for B modifiers were clearly above 5, while the C modifiers were clearly below 5 in both, the typicality and the likelihood rating.<sup>7</sup>

### Comparison to Connolly et al. (2007)

Our main goal was to reproduce the modification effect in Connolly et al. (2007) with a possibility to distinguish between items with and without relevant constraints. For a better comparability, we converted our data to their 1–10 scale. Table 5 shows the descriptive statistics for likelihood and typicality ratings of the four items in comparison to theirs. The two tables show the means and the 0.95 confidence intervals for the probability question and the typicality rating. If the four items are considered together, the ratings resemble Connolly et al.’s result. As we already suspected, the data look quite different if the two relevant and the two non-relevant items are considered separately. The general loss under likelihood for the non-typical modifier is almost solely explained by the data for the relevant items *limousines* and *sofa*. The confidence intervals indicate that the differences from A to C (and also from B to C) are significant for the relevant items but not for the irrelevant ones.

### Relevance Correlations

We showed that the extent of the modification effect was predictable from our modifier relevance assumptions. But to what degree did our assumptions correspond to the subjects’ ratings? And to what degree did their subjective relevance ratings correlate with their individually given likelihoods of the modified statements?

---

<sup>6</sup>Subjects rated having buttons to be more probable than typical for itchy shirts (1.211 [0.013, 2.406],  $p = 0, 048$ ), being long-haired more probable than typical for lambs (0.579 [0.129, 1.029],  $p = 0, 013$ ), being long more probable than typical for inexpensive limousines (1.221 [0.091, 2.330],  $p = 0, 036$ ) and being comfortable less probable than typical for sofas (−0.421, [−0, 825, −0, 017],  $p = 0, 042$ ). Brackets give mean value with 0.95 confidence intervals.

<sup>7</sup>The lowest mean value of a B modifier was 5.92 [5.33, 6.50] in the likelihood rating of *cotton* for *shirt*. The highest mean value of a C modifier was 3.45 [2.70, 4.28] in the likelihood rating of *itchy* for the same item.

**Table 5** Modification effect in comparison to Connolly et al. (2007)

	Shirts, Lamb	Limousines, Sofas	All	Connolly et al. (2007)
(a) Likelihood ratings				
N	37	38	75	400
A	8.30 [7.69, 8.91]	8.37 [7.82, 8.91]	8.33 [7.93, 8.73]	8.38 [8.20, 8.56]
B	7.76 [7.14, 8.39]	7.94 [7.36, 8.52]	7.85 [7.44, 8.27]	7.72 [7.51, 7.93]
C	7.49 [6.69, 8.23]	4.62 [3.75, 5.50]	6.04 [5.37, 6.71]	6.89 [6.66, 7.12]
D	7.42 [6.66, 8.18]	–	–	6.50 [6.27, 6.73]
(b) Typicality ratings				
N	38	38	76	400
A	8.15 [7.54, 8.76]	8.27 [7.65, 8.89]	8.21 [7.79, 8.84]	8.38 [8.20, 8.56]
B	7.77 [7.04, 8.51]	7.80 [7.13, 8.46]	7.79 [7.30, 8.27]	7.72 [7.51, 7.93]
C	7.13 [6.38, 7.89]	4.10 [3.20, 5.00]	5.62 [4.95, 6.29]	6.89 [6.66, 7.12]
D	7.13 [6.36, 7.91]	–	–	6.50 [6.27, 6.73]

**Table 6** Mean judged relevance of non-typical modifiers with 0.95 confidence intervals

Limousines	Sofas	Shirt	Lambs	Lambs (excl)
N = 38	N = 38	N = 38	N = 38	N = 32
3.53	3.37	0.32	3.00	3.56
[2.51, 4.55]	[2.28, 4.48]	[0.07, 0.56]	[1.92, 4.08]	[2.38, 4.75]

The non-typical modifiers were judged to be more relevant for the items *limousines* and *sofas* than for *shirts*. However, for *lambs* people judged origin to be more relevant for the colour than we expected. The item also stood out insofar as many people suspended judgement on this relevance question, while no subject answered “I don’t know” in the relevance rating of any other non-typical modification.<sup>8</sup> Table 6 shows the relevance judgements for the nontypical modifiers. “I don’t know” answers were treated as “0” in the column “Lambs” and excluded in “Lambs (excl)”.

Finally we wanted to know whether the differences in the subjects’ likelihood ratings are related to their individual relevance ratings. We tested this hypothesis for the non-typical modification. First, we calculated the individual modifier effect by subtracting the judgement of the unmodified condition A from the judgement in the modified condition, i.e. C-A.<sup>9</sup> By that means we determined the modification effects for each individual and each item. These values were correlated with the relevance

<sup>8</sup>The differences are probably explained by our research group considering cross value dependencies while the subjects were only confronted with attributes. They might have regarded general evolutionary tendencies that living environments influence appearance, which we did not consider because they are not important for this particular values. We came to the conclusion that it is important to ask for the particular values.

<sup>9</sup>This was possible because we used a within-design.

**Table 7** Kendall's  $\tau$  correlation of relevance and loss in the likelihood rating

Limousines	Sofas	Shirts	Lambs
N = 19	N = 19	N = 19	N = 19
-0.37, $p = 0.046$	-0.38, $p = 0.04$	-0.69, $p < 0.01$	-0.43, $p = 0.03$

ratings of the 19 subjects. This correlation reflects the influence of subjective constraints. Thus we tested whether individually higher or lower modification effects come with individually higher or lower relevance assumptions. It turned out that the status of the modifier was correlated to the modification effect. Kendall's  $\tau$  revealed that subjects with larger decreases in the non-typical modification tended to find the modifier relevant. The correlation and significance is given in Table 7. The correlation is moderate for the items *lamb*, *limousine*, *sofa* and even high for *shirts*, which had a low intersubjective relevance score.<sup>10</sup>

## 4 Conclusion

In this paper, we proposed an extended modification model with constraints. An exploratory study with four of the items used by Connolly et al. (2007) revealed the following tendencies, which largely accord to our assumptions:

- Typicality and likelihood rating  
Likelihood ratings are very similar to typicality ratings. This supports probabilistic approaches to typicality.
- Typical modification  
As already suspected from previous studies, typical modifiers lead to a smaller loss in modification than non-typical ones.
- Valid prior classification for constraints  
The items we suspected to have negative constraints were drastically affected by non-typical modification while the difference to typical modification was negligible for items without constraints.

The prior predictability of loss by non-typical modification has never been investigated so far. However, Connolly et al. (2007) already suspected that the distinctiveness of modification effects is predictable, asserting that “adding *purple* to apple is sure to diminish one’s confidence about its edibility more than adding *ripe* and less than adding *Martian*” (Connolly et al. 2007, p. 14). They take this to be an argument against prototype compositionality. Though prototype compositionality is not

---

<sup>10</sup>In a later larger study with more item, reported in Strößner and Schurz (2020), we were not able to confirm such high correlations between individual modification effects and individual relevance assumptions. However, we were able to confirm that the mean relevance score for a modification is correlated with its mean modification effect.

as straight forward as composing analytic meanings, we disagree with the conclusion that prototypes are not compositional at all. People systematically attribute typical properties to subcategories, even if they are built with meaningless words, as noted by Gagné and Spalding (2014). Their “different but similar” approach, however, would predict that all modifications have roughly the same effect. This is, however, not the case. Our model predicts differences for modifications without constraints and modifications with relevant knowledge constraints. Doubters of prototype theory could argue that the extended modification model includes background beliefs and is thus not about semantics but about belief revision. This argument depends on a very narrow view of compositionality. Understood in the sense of Hampton and Martin (2012) as a process that is not only driven by strict logical intersection but also by common-sense knowledge, the enriched modification model is a model of prototype compositionality.

**Acknowledgements** The research reported in this article was conducted in the project D01 (Frame representation of prototype concepts and prototype-based reasoning) of the SFB991, a collaborative research center that was generously funded by the German Research Foundation (Deutsche Forschungsgemeinschaft). We are grateful to Andrew Connolly for sharing data from his research with Jerry Fodor, Lila Gleitman and Henry Gleitman. Paul Thorn, Sebastian Löbner and an anonymous reviewer provided helpful comments on our manuscript.

## References

- Barsalou, L. W. (1992). Frames, concepts, and fields. In A. Lehrer & E. F. Kittay (Eds.), *Frames, fields, and contrasts*, pp. 21–74. Lawrence Erlbaum Associates Publishers, Hillsdale, New Jersey.
- Connolly, A. C., Fodor, J. A., Gleitman, L. R., & Gleitman, H. (2007). Why stereotypes don't even make good defaults. *Cognition*, 103(1), 1–22.
- Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, 132(2), 163.
- Fodor, J., & Lepore, E. (1996). The red herring and the pet fish: why concepts still can't be prototypes. *Cognition*, 58(2), 253–270.
- Gagné, C. L., & Spalding, T. L. (2011). Inferential processing and meta-knowledge as the bases for property inclusion in combined concepts. *Journal of Memory and Language*, 65(2), 176–192.
- Gagné, C. L. & Spalding, T. L. (2014). Subcategorisation, not uncertainty, drives the modification effect. *Language, Cognition and Neuroscience*, 29(10), 1283–1294.
- Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition*, 15(1), 55–71.
- Hampton, J. A. & Martin, L. J. (2012). Typicality and compositionality: The logic of combining vague concepts. In W. Hinzen, E. Machery, & Werning M. (Eds.), *The Oxford handbook of compositionality*, pp. 385–402. Oxford University Press, Oxford.
- Hampton, J. A., Passanisi, A., & Jönsson, M. L. (2011). The modifier effect and property mutability. *Journal of Memory and Language*, 64(3), 233–248.
- Jönsson, M. L., & Hampton, J. A. (2012). The modifier effect in within-category induction: Default inheritance in complex noun phrases. *Language and Cognitive Processes*, 27(1), 90–116.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.

- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization*, Chap. 2, pp. 27–48. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–605.
- 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.
- Schurz, G. (2012). Prototypes and their composition from an evolutionary point of view. In W. Hinzen, E. Machery, & M. Werning (Eds.), *The Oxford handbook of compositionality*, pp. 530–553. Oxford: Oxford University Press.
- Smith, E. E., Osherson, D. N., Rips, L. J., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive science*, 12(4), 485–527.
- Strößner, C., & Schurz, G. (2020). The role of reasoning and pragmatics in the modifier effect. *Cognitive Science*, 44(2). <https://doi.org/10.1111/cogs.12815>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# A Frame-Theoretic Model of Bayesian Category Learning



Samuel D. Taylor and Peter R. Sutton

**Abstract** Bayesian models of category learning typically assume that the most probable categories are those that group input stimuli together around a maximally optimal number of shared features. One potential weakness of such feature list approaches, however, is that it is unclear how to weight observed features to be more or less diagnostic for a given category. In this theoretically oriented paper, we develop a frame-theoretic model of Bayesian category learning that weights the diagnosticity of observed attribute values in terms of their position within the structure of a frame (formalised as distance from the frame's central node). We argue that there are good grounds to further develop and empirically test frame-based learning models, because they have theoretical advantages over unweighted feature list models, and because frame structures provide a principled means of assigning weights to attribute values without appealing to supervised training data.

**Keywords** Category learning · Bayesian categorisation · Frames · Weighted Naive Bayesian model · Frame-theoretic constraints

## 1 Introduction

Bayesian models of categorisation typically assume that there is both an input to categorisation—the stimulus to be categorised—and an output from categorisation—the (cognitive) behaviour of the categoriser (Kruschke 2008). But in order to count as cognitively adequate, the model must also represent the cognitive processes that mediate between input and output, and take these *representations* to be informative

---

S. D. Taylor (✉)

Institute of Philosophy, Heinrich-Heine-Universität, 40204 Düsseldorf, Germany  
e-mail: [sam.taylor@hhu.de](mailto:sam.taylor@hhu.de)

P. R. Sutton

Institute of Linguistics and Information Science, Heinrich-Heine-Universität,  
40204 Düsseldorf, Germany  
e-mail: [peter.sutton@hhu.de](mailto:peter.sutton@hhu.de)

about the hypothesis space over which Bayesian inference operates. There are a number of possible candidates that could be sourced from cognitive scientific theories—e.g. prototypes, bundles of exemplars, or theory-like structures (Carey 1985; Lakoff 1987; McClelland and Rumelhart 1981; Nosofsky 1988; Rehder 2003). However, it has become standard practice to assume that Bayesian models operate over representations of unstructured lists of features; e.g. feature list representations (Anderson 1991; Sanborn 2006; Goodman et al. 2008; Shafto et al. 2011).

In this paper, we introduce and motivate frames as a candidate for the representations that mediate between (sensory) input and behavioural output, and as the representational format over which Bayesian inference operates in a Bayesian model of category learning. In other words, we introduce frame-theoretic representations (attribute-value structures) as the representational format of the data observed and operated on by the model. Our argument is that the resulting frame-theoretic model of Bayesian category learning is a theoretical improvement on feature list models, because our model can make fine-grained discrimination between competing categories without basing the weighting of attribute values on supervised training data. This is the case because frames—as the representational format of the input to our model—are not mere unordered lists of features, but, rather, are recursive attribute-value structures organised around a central node. For example, instead of three features such as **fur**, **black**, and **soft**, frames represent how these features are related by defining each feature as the value of some attribute i.e., that **fur** has (at least) two attributes COLOUR and TEXTURE, and that the values of these attributes are **black** and **soft**, respectively. As such, frames can be interpreted as assigning attribute values more or less weight depending on properties defined in terms of the structure of frames themselves. As a rough heuristic, our model proposes to weight attribute values as more or less diagnostic depending on whether or not they appear more centrally within a frame. In other words, our model takes a feature’s ‘path distance’ from the central node to determine the diagnosticity of that feature for a given category.

As an example, suppose that the **fur**, **black**, and **soft** values appeared in a frame for a cat. Since, **black** and **soft** are values of attributes of **fur**, and **fur** is the value of an attribute of **cat**, a parameter based on distance from the central node would rank **black** and **soft** lower than **fur**. By incorporating this diagnosticity weighting in our model, we develop a frame-theoretic model of Bayesian category leaning that introduces constraints on the most probable categories in terms of the diagnosticity of the observed features of entities being categorised.

The structure of this paper is as follows. In Sect. 2, we consider weighted Bayesian models of categorisation and argue that there is space to introduce a model that weights the relative diagnosticity of observed features that is not based on labelled training data. Then, in Sect. 3, we introduce a frame-theoretic representation of observed data and categories (e.g. the input and output of a categorisation model), in which frames are recursive attribute-value structures (Barsalou 1992; Barsalou and Hale 1993; Löbner 2014; Petersen 2015; Ziem 2014). Building upon this claim, we argue that the informational-structure of frames can be used to introduce a constraint on the relative diagnosticity of information encoded within a category and/or set of categories, where diagnosticity can be defined partly by properties of frame

structure (distance from the central node). Finally, we outline how feature list models of Bayesian category learning can be extended to operate over frames. On our frame-theoretic approach, the information-structural constraints of the model's frame representational-input influences the conditional probability of possible sets of categories by weighting the diagnosticity of the features of entities being categorised. We consider possible challenges to our model and possible future developments, before concluding that our model is better suited to describe and explain the unsupervised process of categorisation than comparable feature list based alternatives.

## 2 Weighted Bayesian Models of Categorisation

Categorisation is the cognitive process of representing given (natural) domains according to relevant features or properties. These features can be distinguished by our sense modalities—e.g. when we categorise objects in terms of their shape, size, or smell. But these features can also be distinguished by their informational content—e.g. when we can categorise foods in terms of their social role or nutritional content, or when we can categorise animals in terms of their ecological niches or taxonomic group (Shafto et al. 2011). In Bayesian models, categorisation occurs as the result of the model probabilistically grouping together sets of objects with shared features (e.g. **yellow**, **curved**). For instance, in the domain of, say, fruits, **yellow** and **curved** objects will have a relatively higher probability of being categorised together than all **yellow** objects, since other yellow fruits differ widely in their other properties (shape, size etc.), meaning that a clustering of all yellow fruits would yield a category with a below-optimal similarity of features. In this way, Bayesian models of categorisation explain how objects or sets of objects come to be categorised as one type or another (Anderson 1991; Tenenbaum 1999; Fei-Fei and Perona 2005; Wu et al. 2014 amongst many others).

An important question for Bayesian models of categorisation, however, is how models should represent input feature spaces, and, furthermore, how the representation of feature spaces influences the process of Bayesian categorisation. On many approaches to Bayesian category learning, feature inputs are represented as unordered lists of features (Anderson 1991; Sanborn 2006; Goodman et al. 2008; Shafto et al. 2011). And, on this approach, Bayesian categorisation proceeds by making the most probable categories those categories that group input stimuli together around a maximally optimal number of shared features. But, unless weights are added to lists of features in some principled way, this approach can be criticised for failing to provide an account of the relative importance of the features around which categorisation occurs. For example, on this approach the features of **colour**, **shape**, **texture**, **genus**, and **region of first domestication** all count as equally relevant for the differentiation of, say, bananas and oranges. And this seems counter-intuitive, because the representation of certain features—say, **colour** and **shape** in the case of bananas and oranges—appears to be more important for categorisation and so should have a bearing on what is taken to be the maximally optimal grouping of shared features.

In order to resolve the problem of uniformly diagnostic features, weights have been added to Bayesian models of categorisation, which make different features more or less diagnostic for specific categories. Such weighted models, however, face the challenge of finding a principled way to assign weights to individual features. For example, Hall (2007) makes use of a “decision tree-based filter method for setting [feature] weights,” where feature weights are estimated by constructing an unpruned decision tree and looking at the depth at which features are tested in the tree (Hall 2007, p. 121). Similarly, Wu et al. (2014) assign weight values to features by allowing the model to construct an unpruned decision tree that can be used to estimate each feature’s dependence on other features (Wu et al. 2014, pp. 1675–1676). These example models—and many others like them—have contributed to a growing literature that aims to improve the performance of naive Bayesian models while retaining their simplicity and computational efficiency. Notably, however, models which assign weights to features do so on the basis of, for example, frequency of features for categories, where categories are established via supervised learning.

It follows that the weighting schemas implemented by frequency-based approaches are derived from periods of supervised learning; that is, they are schemas that are dependent upon the input of supervised training data (Wu et al. 2014, p. 1676). In principle, there is nothing wrong with the application of such supervised training-based weighting schemas. However, the simplicity and tractability of models based on naive Bayesian assumptions is attractive (Pham 2009), especially if such models can be used in unsupervised learning tasks. This is the challenge that we take up in this paper. We develop a model that maintains the independence assumptions of naive Bayes, whilst assigning weights to features without appealing to weighting schemas derived from a period of supervised learning. The price to pay for this is that one must enrich the data that is input into the model. We do this by taking the input data to be in representational format of frames and not of feature lists. Our justification for this move is set out in Sect. 3, where we argue that there is support for the view that human cognition is structured around richer structures than lists of features and, therefore, that the data made available to learning models ought to be enriched. Furthermore, we argue that the hierarchical structure of frames allows models to assign weights to attribute values in frames.

In the remainder of this paper, we develop a Bayesian frame-based model of category learning. Our model will assign weights to features in virtue of the information structure of the feature spaces observed by the model.<sup>1</sup> In doing so, we drop the assumption that the input feature spaces over which Bayesian models operate are themselves flat and uniformly diagnostic for all categorisation tasks. Our claim is that the relative diagnosticity of features for categories can be captured by enriching the representational format of the data observed by the model. Such an enrichment, we claim, makes explicit how the probability of a system of categories can be cal-

---

<sup>1</sup> Many Bayesian models category learning already presuppose that observed features have an informational structure that makes them more or less diagnostic for a given category, because they introduce certain features—e.g. colour—without making explicit that other features must also be observed; e.g. they introduce the feature **colour** without making explicit that the feature **shape** must also be observed.

culated not only from features (the values of attributes in our terms), but also from the structure of the data itself (such as the path distance that attribute is from the central node). The end result, therefore, is that certain, observed features—e.g. the features **colour** and **shape** in the group of observed features **colour**, **shape**, **texture**, **genus**, and **region of first domestication**—will have more of an influence on the probability of categorising the observed data as one category or another—e.g. as banana or orange.

To be clear, we accept that the evaluation of our model will ultimately be empirical, whereby the model is compared to actual human performance in the course of experimental testing. However, the contribution of this paper is the theoretical development of a model that shows promise as an improvement on current models of Bayesian category learning, since it derives relative feature diagnosticity in an unsupervised manner.

### 3 Frames

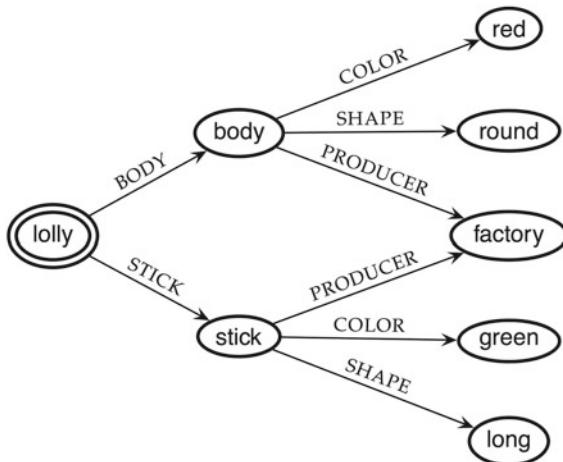
According to Barsalou (1992), frame representations capture the general format of cognition. As attribute-value structures, frames represent both the “general properties or dimensions by which the respective concept is described (e.g., COLOR, SPOKESPERSON, HABITAT …)” and the *values* that each property or dimension takes in any given instantiation “(e.g. [COLOR: red], [SPOKESPERSON: **Ellen Smith**], [HABITAT: **jungle**] …)” (Petersen 2015, p. 151). Thus, “a frame is a representation of a concept for a category which is recursively composed out of attributes of the object to be represented, and the values of these attributes” (Löbner 2014, p. 11). For Barsalou, an attribute is “a concept that describes an aspect of at least some category members”; and values are “subordinate concepts of an attribute” (Barsalou 1992, pp. 30–31). And, thus, a picture emerges of frames as representations of categories that encode, at the attribute level, general properties, dimensions, or aspects of the category in question; and, at the value level, the values taken by specific instantiations of the category in question.

Frames, then, are constituted of attribute-value pairings, where for “every attribute there is the range of values which it can possibly adopt” and “The range of possible values for a given attribute constitutes a space of alternatives” (Löbner 2014, p. 11). For example, an attribute such as COLOUR maps entities to colour values (e.g., [COLOUR: **red**]), and an attribute such as SHAPE maps entities to geometrical values (e.g., [SHAPE: **round**]).<sup>2</sup> Frames can themselves be represented by directed-graphs, whereby labelled nodes specify instantiated regions of the value space and arcs

---

<sup>2</sup> There is an open question about how value spaces are learned by individual subjects. We shall not answer this question here, although we find it plausible that individual subjects have access to value spaces as the result of “hyperpriors” determined by the subject’s biological phylogeny, biological and social ontogeny, and sociocultural embedding (cf. Clark 2015; Newsome 2012).

**Fig. 1** Lolly frame  
(Petersen 2015)



specify attribute designations of regions in the value space (see Fig. 1).<sup>3</sup> Importantly, however, frames cannot be reduced to simple lists of features, because:

[...] it is not possible to simply replace the nodes in the frame definition by their labels, since two distinct nodes of a graph can be labeled with the same type. E.g., we could modify the lolly-frame in [Fig. 1] so that the stick and the body of the described lollies were produced in two distinct factories, where one is located in Belgium and one in Canada. (Petersen 2015, pp. 49–50)

Two questions arise, the answers to which are important for justifying our model:

- (i) Why should we assume that the frames are the representations that mediate between (sensory) input and categorisation of that input (as opposed to feature lists)?;
- (ii) What benefits do frames have as such input over feature lists?

Our simple answer to (i) is that the construction of feature lists implicitly assumes a richer relation between features, which is made explicit when we construct frames. Take the frame in Fig. 1. As a feature list, one could represent part of this information with the following features **has a stick**, **has a body**, **body is red**, **stick is green**. For the latter two in particular, the alternative would be to list two incongruent colour features **red** and **green** (resulting in potential contradiction). Yet, given that features must be more fully specified in this way, such lists of features simultaneously assume an attribute-value structure and make the structure invisible to any model that attempts to form categories on the basis of those features. (Bear in mind, that for a categorisation model, the features **has a stick**, **has a body**, **body is red**, **stick is green** may as well be represented as  $f_1, f_2, f_3, f_4$ , since the fact that two features share ‘stick’ and two features share ‘body’ as part of their labels is not something that a model based on feature lists can access.) Therefore, there is a very real sense in which providing feature lists as data input sells itself short by both implicitly assuming a

<sup>3</sup>Frames can also be represented as attribute-value (AV) matrices (cf. Carpenter 1992; Petersen 2015).

richer structure to the data, but also not allowing any learning model to access that structure.

With respect to (ii), our claim is that the reason why frames are useful and relevant to categorisation is that they can be used to constrain information. In the first place, frames provide constraints on the range of values at any given node, because “information represented in a frame does not depend on the concrete set of nodes. It depends rather on how the nodes are connected by directed arcs and how the nodes and arcs are labelled” (Petersen 2015, p. 49). In other words, if we assume that frames are the category representations that mediate between (sensory) input and behavioural output, then it follows that categories must have a structure that relates the general properties, dimensions, or aspects of a category to the possible values that such general properties, dimensions, or aspects can take. For example, if the value of COLOUR is given as square—e.g. [COLOUR: **square**]—then it is clear that the established ‘category’ is, in fact, no category at all (**square** is not a possible COLOUR value). Thus, it follows that even where a notional ‘category’ contains attribute-value pairs, it may still follow that the ‘category’ in question is impermissible because some of the attributes are assigned infelicitous values.

A second way in which frames constrain information derives from the fact that they are recursive (the value of one attribute can itself have attributes). The central node (graphically, the double-ringed node) indicates what the frame represents (i.e., lollies in the case of Fig. 1). Attribute-value pairs ‘closer’ to the central node encode relatively important, but general, information about the represented object. And attribute-value pairs ‘further’ from the central node encode relatively less important, but more specific, information about the represented object (because they are, e.g., values of attributes of values of attributes of the central node). For example, in Fig. 1 the ‘closer’ attribute-value pairs specify what physical structure and component parts the lolly in question has; and the ‘further’ attributes specify the colour and producer of these components. It follows, therefore, that those attribute-value pairs that are closer to the central node are more likely to be diagnostic of the category into which the object represented should be sorted. Thus, we can conclude that, at least as a rough heuristic, frames with more uniform ‘closer’ attribute-value pairs will represent more likely categories than frames with less uniform ‘closer’ attribute-value pairs (even if the latter has more uniform ‘distant’ attribute-value pairs), because the former categories will be more effective in organising (sensory) input according to more ‘central’ properties.<sup>4</sup> For example, looking again at the lolly frame in Fig. 1, a category containing only red things that may or may not have bodies and sticks will be a less probable category than one which contains objects of different colours that all have bodies and sticks.

In an important paper, Shafto et al. (2011, p. 5) observe that standard approaches to modelling category learning appeal to a ‘single system model’ of categorisation

---

<sup>4</sup>The question of what attribute-value pairs are the most diagnostic for any given (sensory) input or object is an empirical question which we would like to pursue further. Such empirical research is usually undertaken by considering typicality judgements or typicality rankings (Djalal et al. 2016; Rips 1989).

(although the aim of their paper is to develop and motivate a more sophisticated *cross categorisation* model). They define a single system model of categorisation as a model that “embodies two intuitions about category structure in the world: the world tends to be clumpy, with objects clustered into a relatively small number of categories, and objects in the same category tend to have similar features.” So a single system model “assumes as input a matrix of objects and features,  $D$ , where entry  $D_{o,f}$  contains the value of feature  $f$  for object  $o$ ” (Shafto et al. 2011). For the single system model, therefore, “there are an unknown number of categories that underlie the [input],” but the objects that are categorised within the same category “tend to have the same value for a given feature” (Shafto et al. 2011). As a result, the ultimate goal of the model is to infer—by means of establishing groupings within  $D$  according to shared features—likely set of categories,  $w \in W$ , where the process of categorisation occurs as the result of a trade-off between two *goals* or *constraints*: “minimizing the number of [categories] posited and maximizing the relative similarity of objects within [each category]” (Shafto et al. 2011).

Such models, and the model we develop here, make independence assumptions regarding feature spaces (value spaces for attributes, in our terms). For example, that the colour of the body of a lolly is independent from the manufacturer of the body. Single system models of categorisation proceed by partitioning the hypothesis space—e.g. the objects in the input matrix,  $D$ —according to more or less probable sets of categories,  $w$ . Finally, the posterior probability of hypotheses given the data ( $p(w|D)$ ) is calculated, where this posterior probability is influenced by the extent to which objects grouped into categories share features (are homogeneous) (Shafto et al. 2011, p. 6).

Replacing feature lists with frames amounts to making the input matrix  $D$  richer. When the input matrix specifies frames and not merely feature lists, the structure of frames can be used to define parameters for a categorisation model. Here, we investigate the possibility of exploiting the fact that frames are hierarchical. Graphically, each node can be measured in terms of path distance from the central node. Added to the fact that attributes are functional, this allows us to define, as a rough heuristic, the relative diagnostic strength of an attribute value from that value’s distance from the central node. Hence, by including in  $D$  weighted values, where weights are derived from frame structure, Bayesian inference operates over a richer information set.

Consider the simple feature list matrix for four witnessed objects  $a, b, c, d$  and four features **fur**, **feathers**, **brown**, **black** in Table 1. If we assume that, even as feature lists, these features can be grouped into classes, which we label colour and layer, the joint probability distribution for the data can be given as shown in Table 2.

The possible groupings of objects into categories for this sample already numbers 15. Four such are given in (1) with the additional information of how these groupings relate to the features of objects.

$$\begin{aligned}
w_1 &= \begin{cases} \mathbf{fu} \wedge \mathbf{br} = \{a\} \\ \mathbf{fu} \wedge \mathbf{bl} = \{b\} \\ \mathbf{fe} \wedge \mathbf{br} = \{c\} \\ \mathbf{fe} \wedge \mathbf{bl} = \{d\} \end{cases} & w_2 &= \begin{cases} \mathbf{fu} = \{a, b\} \\ \mathbf{fe} \wedge \mathbf{br} = \{c\} \\ \mathbf{fe} \wedge \mathbf{bl} = \{d\} \end{cases} \\
w_8 &= \begin{cases} \mathbf{fu} = \{a, b\} \\ \mathbf{fe} = \{c, d\} \end{cases} & w_{15} &= \{\mathbf{fu} \vee \mathbf{fe} \vee \mathbf{br} \vee \mathbf{bl} = \{a, b, c, d\}\}
\end{aligned} \tag{1}$$

However, the number of possible sets of categories increases exponentially with the number of objects. This presents a categorisation challenge. Given a huge number of hypotheses for categorising a set of objects, the options must be whittled down. Bayesian approaches to categorisation can do this by calculating the maximum probability for some set of categories  $w_i$ , given the data  $D$ , namely:  $\text{MAX}_{w_i \in W}[p(w_i | D)]$  (such that these probabilities can be updated in the light of new data). (Other alternatives include Markov Chain Monte Carlo Variational Bayesian methods.) For example, Shafto et al. (2011), following Anderson (1991), argue that this probability depends on the prior probability of assigning objects to categories (in a set of categories  $w$ ) and the probability of the data given a set of categories.

We adopt Shafto et al.'s (2011) use of two parameters and the way in which they contribute to calculating  $p(w | D, \alpha, \delta)$ <sup>5</sup>:

$$p(w | D, \alpha, \delta) \propto p(w | \alpha) \times p(D | w, \delta) \tag{2}$$

In (2),  $p(w | \alpha)$  contains the parameter  $\alpha$  which sets the extent to which the number categories should be minimised.  $p(D | w, \delta)$  contains the parameter  $\delta$  which sets the extent to which features of objects within categories should be similar (i.e., that members of categories should have the same feature/attribute values).

**Table 1** Distribution of skin covering and colour features (simulated)

	a	b	c	d
<b>fur</b>	1	1	0	0
<b>feathers</b>	0	0	1	1
<b>brown</b>	1	0	1	0
<b>black</b>	0	1	0	1

**Table 2** Joint probability distribution:  $f_{L,C}(l, c)$

		LAYER ( $L$ )	
		<b>fur</b>	<b>feathers</b>
<b>COLOUR (<math>C</math>)</b>	<b>brown</b>	0.25	0.25
	<b>black</b>	0.25	0.25

<sup>5</sup>Our model differs from theirs, however. See Appendix 1.

As a simple example of how these parameters work, take the data in Table 1. If the  $\alpha$  parameter is set to maximally minimise the number of categories, then maximising  $p(w|\alpha)$  would select  $w_{15}$  in (1); namely, a singleton set of one category that includes all objects so far observed. If, however, the parameter  $\delta$  is set to maximise feature harmony within categories, then maximising  $p(D|w, \delta)$  would select  $w_1$  in (1); namely, a set of categories that contains as many categories as there are ways to distinguish objects by their features.

Such feature list models have been implemented for categorisation tasks (Chater and Oaksford 2008; Shafto et al. 2011). However, notice that for some data sets, although we would intuitively categorise some entities together, unweighted feature lists provide an insufficient amount of information to distinguish between competing hypotheses. Take, once more, the data in Table 1. No matter how one sets parameters such as  $\alpha$  and  $\delta$  in a feature list based Bayesian categorisation model, the probability value for  $w_8$  in (1) could not differ from the value for  $w_9$  in (3):

$$w_9 = \begin{cases} \mathbf{br} = \{a, c\} \\ \mathbf{bl} = \{b, d\} \end{cases} \quad (3)$$

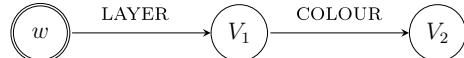
The reason for this is because even if we grant that a model can be set up to see brown versus black and feathers versus fur as two distinct comparison classes, the flat nature of feature lists does not allow for (observed) relations between features to be expressed, which, were they articulated, could be used to inform judgements regarding probable sets of categories. In other words, as has been recognised, feature lists must, at the very least, be weighted in some principled way. The problem is that, in an unsupervised learning task, it is difficult to justify the selection of one feature over another.

Given frames as input data, however, such weightings can be defined by parameterising the structure of frames themselves. In other words, with frames, a categorisation model can be defined that can distinguish cases such as  $w_8$  and  $w_9$ . This is made possible because frames introduce a hierarchy between feature values in virtue of the fact that some values are values of attributes of other values. For the case in hand, for example, **black** and **brown** could be observed to be values of a COLOUR attribute, such that COLOUR is an attribute of the values **fe** and/or **fu**.<sup>6</sup> That is to say the data in Table 1 could license the attribute-value structure shown in Fig. 2.

---

<sup>6</sup>In this paper, we are making the assumption that fur/feather-based categories are preferable. We take this to be reasonable on common-sense grounds. However, we also accept that there may be cross-cultural variation in the kinds of feature-based categories preferred. For example, it may be the case that individuals in certain cultures—e.g. Yucatec-speaking cultures—prefer material-based categories, while individuals in other cultures—e.g. English-speaking cultures—prefer shape-based categories (Lucy and Gaskins 2001). The kinds of cross-cultural differences that may be apparent in categorisation tasks cannot be dealt with adequately in this paper due to lack of space. Still, it is worth noting that our model—like any other Bayesian model of category learning—could be supplemented with further constraints to account for such differences in categorisation tasks. Such supplementation would first have to be justified in the light of ongoing debates about the relation between language, culture, and thought (cf. McWhorter 2014; Lucy 1992a,b).

**Fig. 2** Attribute-value structure for data in Table 1



**Table 3** Distribution of fur layer and colour features relative to distance (simulated)

	a	b	c	d
$\langle \text{fur}, 1 \rangle$	1	1	0	0
$\langle \text{feathers}, 1 \rangle$	0	0	1	1
$\langle \text{brown}, 2 \rangle$	1	0	1	0
$\langle \text{black}, 2 \rangle$	0	1	0	1

Our proposal is that, in general, the importance of the similarity of feature values of objects within categories is proportional to how ‘close’ these feature values are to the central node measured by (minimum) path distance. The intuitive idea is that properties of objects within the same category tend to be similar, at least in terms of type, when these properties are more diagnostic of the category in question (see Sect. 3). Take the frame from Petersen (2015) in Fig. 1. The type of value for the BODY and STICK attributes will be very similar across different lollies. Indeed, if something had, e.g., lolly properties but no stick, one might judge it to be a sweet, not a lolly. However, the shape, colour, and producer for each lolly component may vary to a greater extent without giving one cause to judge, e.g., that two differently coloured objects belong to different categories qua *lolly* or *not a lolly*.

Using unweighted feature lists alone, one cannot formally capture that similarity between values is more important for more central nodes. With frames we can. Given that we will not here be exploiting further properties of frames, data sets can be minimally changed to include a distance measure. For the frame in Fig. 2, for example,  $V_1$  measures a distance of 1 from the central node.  $V_2$  measures a distance of 2. (For more complex frames, this means that there may be multiple values that measure the same distance.<sup>7</sup>) This requires a fairly minimal adjustment in how data sets are represented. The data in Table 1, for example, will be represented as in Table 3. The adjustment made is that we now represent features as pairs  $\langle f, d \rangle$  where  $f$  is a feature (e.g. **brown** or **feathers**) and  $d$  is a measure of distance such that  $d \in \mathbb{N}$ . This change is not trivial. Enriching the data set could be seen as some kind of cheat, i.e., by providing more information that guides the process of forming categories. However, as we argued in Sect. 3, such structure is often implicit in feature lists, even if it invisible to the learning model. In our model, we make this implicit information available.<sup>8</sup>

<sup>7</sup>We assume, in cases where a node is connected to the central node along multiple paths, that this is calculated as the minimum distance.

<sup>8</sup>It should be stressed that we lose a lot of information by compressing frames in this way. However, we do this for simplicity and do not rule out that retaining more information in frames may be required in future developments of this model.

A full specification of our model is given in Appendix 1. In brief, we calculate the value for  $p(w|\alpha)$  from the sum of the entropy of the set of categories in  $w$  with respect to the assignment of objects to categories in  $w$ , weighted by  $\alpha$ . In other words, in terms of the average amount of information required to determine which object a category is in, given a set of categories. A  $w$  with only one category will minimise entropy (no information is required to know which category an object is in because all objects are in one category). This translates into a high value for  $p(w|\alpha)$ . Depending on the value of  $\alpha$ , a  $w$  with many categories will have comparably higher entropy (especially if the categories are evenly distributed/of similar size). This translates into a comparably lower value for  $p(w|\alpha)$ . Values of  $p(D|w, \delta)$  are calculated from the  $\delta$ -weighted entropy of each category with respect to the features of objects within that category. If all objects within each category have the same features, then entropy will be minimised (one would need no information to know which features an object has given the category it is in). This translates into a high value for  $p(D|w, \delta)$ . If objects in the same category differ with respect to their attribute values, then, depending on the setting for  $\delta$ , this probability will be lower.

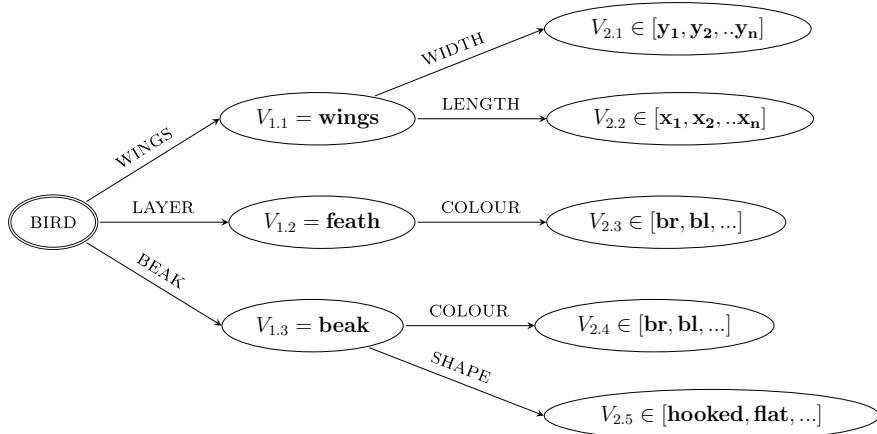
The difference between our model and one based on feature lists, therefore, is that unsupervised feature list models do not have a principled way to weight similarity with respect to some features more heavily than similarity with respect to others. For feature list models, given the data set in Fig. 1 and  $w_8$  and  $w_9$  from (1) and (3), for example,  $p(w_8|D, \alpha, \delta) = p(w_9|D, \alpha, \delta)$  for all settings of  $\alpha$  and  $\delta$ . However, our frame-based model can discriminate between these two sets of categories. Objects in categories in  $w_8$  have the same attribute values at distance 1 from the central node (viz. **fe** and **fu**), but different attribute values at distance 2 from the central node (viz. **br** and **bl**). In contrast, objects in categories in  $w_9$  have different attribute values at distance 1 from the central node (viz. **fe** and **fu**), and the same attribute values at distance 2 from the central node (viz. **br** and **bl**). (See Appendix 1 for details.)<sup>9</sup>

### 3.1 Challenges and Future Developments

**Refining the model to discriminate between subkinds/superkinds.** This kind of model opens up an intriguing avenue for further research: we could define levels of granularity for categorisation by manipulating the function which underpins  $\delta$ . For example, relatively coarse-grained categorisation would prioritise similarity of object features only for nodes that are small distances from the central node. This

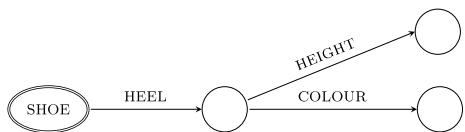
---

<sup>9</sup>We do not claim that there is no other way to do this. For example, possible sets of categories, formed from unweighted feature list input, could be ranked according to other principles such as *simplicity* in which sets of categories are preferred if they minimise similarities within categories and maximise differences between categories (Chater 1999; Pothos and Chater 2002). Indeed, it is an open and interesting question whether our model ends up approximating the results of a simplicity-driven strategy, or, if not, whether both a frame based input and a simplicity-driven categorisation strategy could be combined in some way. We leave the comparison between our model and others for future work.



**Fig. 3** (Partial) Frame for BIRD

**Fig. 4** Partial frame for SHOE



might, for example, cluster birds together and mammals together. If, however,  $\delta$  is set to push towards similarity of values in ‘further out’ nodes, then distinctions between categories would be more fine grained. This could, for example, allow for the BIRD category to be partitioned into species of birds. The reason for this is that there is a general tendency for birds to be similar with respect to values closer to the central node (e.g. **feathers**, **wings**, **beak** etc.), but dissimilar with respect to less central values. For example, beaks, wings, and feathers may differ with respect to shape, size, and colour. The basic idea is shown in Fig. 3. If values at distance 1 from the central node are enforced to be similar ( $V_{1.1}$ ,  $V_{1.2}$ , and  $V_{1.3}$ ), but values at distance 2 can differ ( $V_{2.1}$ – $V_{2.5}$ ), then we would expect birds to be categorised together. However, if the setting for  $\delta$  was such that values at distance 1 and at distance 2 were enforced to be (more-or-less) similar, we would get a categorisation of, say, different bird species.

An interesting avenue for further research is whether or not our model, which is a single system model in the sense of Shafto et al. (2011), could be used as a cross categorisation model by manipulating the function that underpins the  $\delta$  parameter.

**Distance may be insufficient as a measure.** Our model has limitations as a result of our simplistic adoption of distance from the central node as the basis for justifying the weighting of certain attribute values over others, namely, for some cases, such a coarse measure is unlikely to get the right results. For example, take a frame for shoes such that one wishes to discriminate high-heeled shoes from loafers. In such a case the height of the heel is surely a highly diagnostic factor. However, as indicated

by Fig. 4, other, far less relevant factors, such as the colour of the heel will appear at the same distance from the central node. Developments of our account will therefore have to investigate if there are other features of frames that can be parameterized in a categorisation model to capture such cases. For example, an extra feature of frames that we have not discussed here are constraints between values. For example, finding out the height of a shoe's heel may be highly informative as to other attribute values (such as the shape of the upper, the (un)likelihood of shoelaces etc.). One possible extension would therefore be to enrich the model with a parameter based upon numbers of constraints a node has linking it to other nodes. (The colour of a heel will be less likely to constrain other values than the height of the heel.)<sup>10</sup>

**Necessity of empirical verification of the model.** We submit that our frame-theoretic model of Bayesian category learning is an important theoretical development in one crucial respect: the model incorporates weights on the relative diagnosticity of attribute-value pairs without having to index such weightings to properties discerned from a period of supervised learning. In other words, our model provides an unsupervised way of introducing weights on the relative diagnosticity of attribute-value pairs, such that one need not train the model on a data set already imbued with category distinctions. However, we also accept that, in this paper, we have only been able to make explicit a *theoretical* difference between our model and comparable alternatives. It follows that our model—if it is to be taken as an accurate representation of human performance in categorisation tasks—must be empirically tested. That is, experimental methods must be employed to compare the categorisation performance of our model with the categorisation performance of other available models. In this way, our model must be comparatively evaluated according to how well it accounts for a given set of data relating to human performance, so that it can be empirically demonstrated that our model better explains human performance than its rivals. We therefore plan to test our model empirically in future research.

## 4 Conclusion

Although a number of representational formats have been exploited to account for the input to Bayesian categorisation models, it remains unclear which is best suited to modelling human categorisation. On the received view, Bayesian inference is taken to operate over input in the form of object-feature list matrices. Although such models have made progress, we have argued here that they only have sufficient discriminatory power because they tend to implement weighting schemas based on supervised learning (weights are derived from exemplars of categories provided in a period of supervised (or semi-supervised) learning).

Our central contribution has been to introduce and exploit frames as the representational format of the input to Bayesian models of category learning. Frames have a richer informational-structure than do feature lists, and so can be used to determine

---

<sup>10</sup>Such an enrichment would amount to dropping many of our independence assumptions, however.

the weighted diagnosticity of the information encoded within a category. As a result, the frame-based model we developed can discriminate between competing sets of categories without having to define weights based on samples of data labelled with categories. In other words, we have given a theoretical basis for a Bayesian categorisation model that, in principle, can approximate weighted naive Bayesian models without a period of supervised learning or weakening the independence assumptions of such models. This follows because the structure frames inherently have (and feature lists lack) can be used to define such weights directly from training data that is not tagged with categories to be learned.

Our adoption of frames as representations of data input and category output extends and consolidates the enlightened Bayesian paradigm, which looks to developments in the cognitive sciences to inform Bayesian modelling techniques (Chater et al. 2011; Jones and Love 2011). As postulates of cognitive scientific theories, frames are already a well-established representational architecture (among many others, see Barsalou 1992; Löbner 2014; Ziem 2014). However, until now, the theoretical benefits of frames had not been made explicit within the context of Bayesian models of category learning. By arguing that frames allow for the development of a more intuitively discriminatory model of category learning based on enriched input, we hope to have shown one way that an account of categorisation based upon the mathematical ideals of Bayesianism can still be subject to principled representational constraints. Although we accept that more work is needed to spell-out the evolutionary and practical relationship between Bayesian inference and (mental) representations in the broader domain of cognitive development, we think that our frame-theoretic approach to Bayesian category learning serves as a welcome further step on the path to developing a mechanistically-grounded and formally rigorous picture of cognition.

**Acknowledgements** We would like to thank the participants of *Cognitive Structures: Linguistic, Philosophical and Psychological Perspectives* (2016) for their constructive comments and critique. We would also like to extend our thanks to the two anonymous reviewers for their helpful recommendations and advice about how the paper could be improved. This work was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft) as part of the Collaborative Research Centre 991: The Structure of Representations in Language, Cognition, and Science, projects C09 and D02. Thanks also to the members of C09 and D02 for their support.

## Appendix 1: A Frame-Based Bayesian Categorisation Model

Our model is based, like other single system models, on the calculation of  $p(w|D, \alpha, \delta)$  from the joint probability distribution over  $w$ ,  $D$ ,  $\alpha$ , and  $\delta$  (elements of the model). We use the same formula (reprinted here with an  $M$  label on  $p$  to indicate the probability function based on this joint distribution):

$$p_M(w_i|D, \alpha, \delta) \propto p_M(w_i|\alpha) \times p_M(D|w_i, \delta) \quad (4)$$

**Table 4** Definitions for elements of the frame based Bayesian categorisation model

$O = \{o_1, \dots, o_n\}$	A set of observed objects
$C = \{c \mid c \neq \emptyset, c \subseteq O\}$	A set of categories
$W = \{w \mid w \neq \emptyset, w \subseteq C\}$	A set of sets of categories
$F = \{\mathbf{f} \mid \mathbf{f} \subseteq O\}$	A set of attribute values (i.e. a set of predicates of objects)
$\iota = f : F \rightarrow \mathbb{N}_{>0}$	A function from attribute values to their distance from the central node
$A = \{(\mathbf{f}, n) \mid \mathbf{f} \in F, n = \iota(\mathbf{f})\}$	A set feature-distance pairs (i.e. a set of distance indexed features)
$D = \{(o, X) \mid o \in O, X \subseteq A\}$	The data: a set of tuples such that for each object, there is a set of feature-distance pairs
$\alpha \in [0, 1]$	The small categories preference parameter
$\delta = f : \mathbb{N}_{>0} \rightarrow \mathbb{R}$	The similar features preference parameter. A function from distance measures to real numbers

We maintain the small categories preference parameter  $\alpha$ , but the similar features preference  $\delta$ , on our model, sets the preference for how strongly distance from the central node affects the overall similarity score for a set of categories. Definitions of elements of the model are given in Table 4. Categories are sets of objects and category schemas are sets of categories. The data input for the model consists of frames, here simplified to objects paired with attribute values and a distance of this value from the central node. Distance from the central node forms the basis for the weighting of attribute values determined by  $\delta$ .

We assume, for simplicity, that for any set of categories,  $w$ , no object is in more than one category and every object is in a category. (Sets of categories completely partition the domain of objects.) In other words, as given in (5), for a set of objects,  $O$ , for each  $w$ , we have a distribution over the categories  $c_i \in w$  (the probability function is accordingly labelled  $O, w$ , we suppress  $O$  in most of the following since we will not consider cases for multiple  $O$  sets).

$$\sum_{c_i \in w} p_{O,w}(c_i) = 1 \quad (5)$$

The prior probability of a category  $c$  relative to a set of categories  $w$  is calculated as the number of objects in the category divided by the number of objects so far observed:

$$\text{For each } c \in w, \quad p_{O,w}(c) = \frac{|c|_w}{|O|} \quad (6)$$

Other distributions occur at the level of nodes in frames. Each node has a set of possible values (e.g., **red**, **green** etc., for COLOUR, and **feathers**, **fur**, **scales** etc. for COVERING). We say more about such distributions in Appendix 1.2.

**Table 5** The effect of  $\alpha$  on calculating the prior  $p(w|\alpha)$  for the data in Table 3 restricted to  $w_1$  and  $w_{15}$ 

		$\alpha$	1	0.5	0.1	0
$p(w_1 \alpha)$		0.20	0.3	0.47	0.50	
$p(w_{15} \alpha)$		0.80	0.6	0.53	0.50	

## 1.1 The $\alpha$ Parameter

The intuitive idea behind the calculation of  $p_M(w|\alpha)$  is that  $w$  should minimise entropy over the object space (minimise the average amount of information required to identify in which category in  $w$  an object belongs). This is given in (7). If alpha is set to 1, then the probability is proportional to the inverse log of the entropy of  $w$ . If  $\alpha = 0$ , then, assuming a base-2 logarithm, for all  $w \in W$ ,  $p_M(w|\alpha) \propto 2^0$  (i.e.  $\propto 1$ ), thus all  $w \in W$  would receive the same prior.<sup>11</sup> In other words, there would be no preferential effect of reducing (or increasing) the number of categories.

$$\text{For all } w \in W : p_M(w|\alpha) \propto 2^{\alpha} \left( \alpha \times \sum_{c_i \in w} (p_w(c_i) \times \log_2(p_w(c_i))) \right) \quad (7)$$

As an example of how  $\alpha$  operates, consider four objects  $a, b, c, d$  and a space of two category sets  $w_1, w_{15}$ . If  $w_1 = \{c_1 = a, c_2 = b, c_3 = c, c_4 = d\}$  and  $w_{15} = \{c_5 = \{a, b, c, d\}\}$ , then, for varying values for  $\alpha$ , we get the results in Table 5 (values given to 2 decimal places).

## 1.2 The $\delta$ Parameter

The intuitive idea behind the calculation of  $p(D|w_i, \delta)$  is that, with respect to the values for an attribute, each category should minimise entropy (weighted by distance the attribute is from the central node). In other words, minimise the average amount of information it takes to decide which properties an object has if it is in a particular category.

Given that each  $d \in D$  is a tuple of an object and a set of attribute value-distance pairs, calculating  $p_M(D|w, \delta)$  turns on calculating, for each category  $c$  in  $w$ , the probability that the objects in  $c$  have some particular value for the relevant attribute. Let  $|\mathbf{f}_j|_{c_k, w, D}$  be the number of times the attribute value  $\mathbf{f}_j$  occurs as a value in category  $c_k \in w$  for a data set  $D$ . Let  $|c_k|_{w, D}$  be the number of objects in  $c_k \in w$ .  $p_{w,D}(\mathbf{f}_j|c_k)$  is, then:

---

<sup>11</sup>The actual probability is calculated by dividing by the sum of the values given in (7) over all  $w \in W$ .

$$p_{w,D}(\mathbf{f}_j|c_k) = \frac{|\mathbf{f}_j|_{c_k,w,D}}{|c_k|_{w,D}} \quad (8)$$

namely, for a set of categories  $w$ , the total number of times objects in  $c_k \in w$  have value  $\mathbf{f}_j$ , divided by the total number of objects in  $c_k$ . This forms a distribution for any set of attribute values that are the mutually exclusive values of some attribute (e.g., a distribution over **feathers** and **fur**, and a distribution over **black** and **brown** in our toy example).

The entropy values for attribute value spaces, given a category, are weighted depending on the distance  $d$  the feature is from the central node. This weighting is set by  $\delta$ , which is a function from  $d$  to a real number in the range  $[0, 1]$ . The weighted entropy value for a category is, then, the sum of the weighted sum of the surprisal values for each attribute value, given a category, also weighted by  $\delta$ . The weighted entropy value for a set of categories  $w$  is the weighted average of the entropy values for each category in  $w$  (relative to  $p_w(c)$ ). So, for all  $w \in W$ :

$$p_M(D|w, \delta) = 2^{\wedge} \left( \sum_{c_k \in w} p_w(c_k) \times \sum_{(\mathbf{f}_j, n_j) \in \pi-2(D)} \left( p_{w,f,c}(\mathbf{f}_j|c_k) \times \log_2(p_{w,f,c}(\mathbf{f}_j|c_k)) \times \delta(n_j) \right) \right) \quad (9)$$

Intuitively,  $p_M(D|w, \delta)$  is a measure on how well the data is predicted by each  $w$  (weighted by  $\delta$ ). This value will be 1 if every piece of data (an object and its attribute values and distances) falls into a totally homogenous category with respect to the objects it contains. This is because the average amount of information to determine the attribute values of members of each category is 0. As categories get more and more heterogeneous, the value of  $p(D|w, \delta)$  will get lower. This is because the average amount of information need to determine the attribute values of members of each category is high.

For example, for the data in Table 3, so with four objects  $a, b, c, d$ , and also with the four category sets  $w_1, w_8, w_9, w_{15}$ , if  $w_1 = \{c_1 = \{a\}, c_2 = \{b\}, c_3 = \{c\}, c_4 = \{d\}\}$ ,  $w_8 = \{c_5 = \{a, b\}, c_6 = \{b, c\}\}$ ,  $w_9 = \{c_7 = \{a, c\}, c_8 = \{b, d\}\}$ , and  $w_{15} = \{c_9 = \{a, b, c, d\}\}$  and attribute values are as displayed in Table 3, then we get the impact of altering the  $\delta$  function as given in Table 6 (values given to 2 decimal places). Since  $w_1$  contains only singleton categories, the probability of the data given  $w_1$

**Table 6** The effect of  $\delta$  on calculating the likelihood  $p(D|w, \delta)$  for the data in Table 3 for  $w_1, w_8, w_9$  and  $w_{15}$ , for  $\delta(n_j) = n_j^0$ ,  $\delta(n_j) = n_j^{-1}$ , and  $\delta(n_j) = n_j^{-2}$

	$\delta(n_j)$	$n_j^0$	$n_j^{-1}$	$n_j^{-2}$
$p(D w_1, \delta)$		1	1	1
$p(D w_8, \delta)$		0.5	0.71	0.84
$p(D w_9, \delta)$		0.5	0.5	0.5
$p(D w_{15}, \delta)$		0.25	0.35	0.42

is 1 no matter how  $\delta(n_j)$  is defined, since for all attribute values and all categories  $P_{w_1, f, c}(\mathbf{f}_j | c)$  equals 1 or zero (so the weighted entropy value is 0 and  $2^0 = 1$ ). The worst performing is  $w_{15}$ , since this contains only one category so heterogeneity for features is high (this is mitigated a little when  $\delta(n_j)$  is defined to decrease the homogeneity requirement for attribute values with larger distances from the central node).

We now turn to the comparison between  $w_8$  and  $w_9$  (which is important for our toy example). In the case where  $\delta(n_j) = n_j^0$  (i.e. where  $\delta(n_j)$  is always equal to 1), there is no weighting towards the importance of similarity of values with respect to being close to the central node. This gives us the same result as would be given for a simple unweighted feature list. In other words, given some things that are furry and black, furry and brown, feathered and black, and feathered and brown, the model has no preference towards grouping furry things together and feathered things together over grouping black things together and brown things together. When  $\delta(n_j) = n_j^{-1}$ , the result is that entropy is weighted to be halved for values at a distance of two nodes away from the central node. When  $\delta(n_j) = n_j^{-2}$ , the result is that entropy is weighted to be quartered for values at a distance of two nodes away from the central node. This translates into an increasing preference for no entropy at the inner most nodes and an allowance of higher entropy at further out nodes.

## References

- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological review*, 98(3), 409.
- Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In E. Kittay & A. Lehrer (Eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization* (pp. 21–74). Erlbaum.
- Barsalou, L. W., & Hale, C. (1993). Frames, concepts, and conceptual fields. In I. Van Mechelen, J. Hampton, R. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 97–144). Academic Press.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.
- Carpenter, B. (1992). *The logic of typed feature structures*. Cambridge: Cambridge University Press.
- Chater, N. (1999). The search for simplicity: A fundamental cognitive principle. *Quarterly Journal of Experimental Psychology: Section A*, 52(2), 273–302.
- Chater, N., & Oaksford, M. (2008). The probabilistic mind: Where next? In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 501–514). Oxford: Oxford University Press.
- Chater, N., Goodman, N., Griffiths, T. L., Kemp, C., Oaksford, M., & Tenenbaum, J. B. (2011). The imaginary fundamentalists: The unshocking truth about Bayesian cognitive science. *Behavioral and Brain Sciences*, 34(4), 194–196.
- Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford: Oxford University Press.
- Djalal, F. M., Ameel, E., & Storms, G. (2016). The typicality ranking task: A new method to derive typicality judgments from children. *PLoS ONE*, 6(11), 1–17.
- Fei-Fei, L., & Perona, P. (2005). A Bayesian hierarchical model for learning natural scene categories. In *IEEE Computer Society Conference, Computer Vision and Pattern Recognition, 2005, CVPR 2005* (Vol. 2, pp. 524–531).

- Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, 32(1), 108–154.
- Hall, M. (2007). A decision tree-based attribute weighting filter for naive Bayes. *Knowledge-Based Systems*, 20(2), 120–126.
- Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34(4), 169–188.
- Kruschke, J. K. (2008). Bayesian approaches to associative learning: From passive to active learning. *Learning and Behavior*, 36(3), 210–226.
- Lakoff, G. (1987). Cognitive models and prototype theory. In U. Neisser (Ed.), *Concepts and conceptual development* (pp. 63–100). Cambridge: Cambridge University Press.
- Löbner, S. (2014). Evidence for frames from human language. In *Frames and concept types* (pp. 23–67). Springer International Publishing.
- Lucy, J. A. (1992a). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.
- Lucy, J. A. (1992b). *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.
- Lucy, J. A., & Gaskins, S. (2001). Grammatical categories and the development of classification preferences: A comparative approach. In *Language acquisition and conceptual development* (pp. 257–283).
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1 an account of basic findings. *Psychological Review*, 88, 375–407.
- McWhorter, J. H. (2014). *The language hoax: Why the world looks the same in any language*. Oxford University Press.
- Newsome, W. (2012). Complementing predictive coding. *Frontiers in Psychology*, 3, 554.
- Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 700–708.
- Petersen, W. (2015). Representation of concepts as frames. In T. Gamerschlag, D. Gerland, R. Osswald, & W. Petersen (Eds.), *Meaning, frames, and conceptual representation* (pp. 39–63). Düsseldorf: Düsseldorf University Press.
- Pham, D., & Ruz, G. (2009). Unsupervised training of Bayesian networks for data clustering. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* (Vol. 465, No. 2109, pp. 2927–2948).
- Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorisation. *Cognitive Science*, 26(3), 303–343.
- Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1141–1159.
- Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). Cambridge: Cambridge University Press.
- Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2006). A more rational model of categorization. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 726–731).
- Shafto, P., Kemp, C., Mansinghka, V., & Tenenbaum, J. B. (2011). A probabilistic model of cross-categorization. *Cognition*, 120(1), 1–25.
- Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. In M. S. Kearns, S. A. Solla, & D. A. Cohn (Eds.), *Advances in neural information processing systems* (Vol. 11, pp. 59–65). Cambridge, MA: MIT Press.
- Wu, J., Pan, S., Cai, Z., Zhu, X., & Zhang, C. (2014). Dual instance and attribute weighting for naive Bayes classification. In *2014 International Joint Conference on Neural Networks (IJCNN)* (pp. 1675–1679).
- Ziem, A. (2014). FrameNet, Barsalou frames and the case of associative anaphora. In *Studies in language and cognition* (pp. 93–112).

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Extremes are Typical. A Game Theoretical Derivation



Robert van Rooij and Thomas Brochhagen

**Abstract** In this paper we argue that a typical member of a class, or category, is an extreme, rather than a central, member of this category. Making use of a formal notion of representativeness, we can say that a typical member of a category is a stereotype of this category. In the second part of the paper we show that this account of typicality can be given a rational motivation by providing a game-theoretical derivation.

**Keywords** Typicality · Representativeness · Extreme · Game theory

## 1 Typicality: Prototypes Versus Stereotypes

In cognitive psychology, a typical representative of class  $X$  is normally called its *prototype*. At least since the work of Rosch (1973) in psychology, a prototype of a category is standardly seen as an item that is most similar to all other members of the category: a *central member* of the category. It is standardly assumed that category membership is a graded affair, and that goodness-of-exemplar judgments depend on similarity to the prototype.

But is a typical member of a category really a central member of this category? A simple Google search seems to question this view. The man that comes up very prominently when one does a simple Google search of a typical, or real man is Rambo. Whatever one can say of Rambo, he is *not an average man*. Very similar pictures of real *tall* men, and real *scientists* give rise to similar conclusions.

Our Google search should obviously not be taken too seriously, but it is in line with many experimental findings in cognitive psychology of what we think of typical exemplars. First, Hampton (1981) found that at least for abstract categories, central

---

R. van Rooij (✉) · T. Brochhagen  
Institute for Logic, Language and Computation University of Amsterdam,  
Amsterdam, Netherlands  
e-mail: [r.a.m.vanrooij@uva.nl](mailto:r.a.m.vanrooij@uva.nl)

T. Brochhagen  
e-mail: [thomasbrochhagen@gmail.com](mailto:thomasbrochhagen@gmail.com)

tendencies are not a good predictor of goodness-of-exemplar judgments. Second, Barsalou (1985) showed that *ideals*, rather than central exemplars, are better determinants of category goodness in goal-based categories such as ‘foods to eat on a diet’ (food with zero calories) and ‘ways to hide from the Mafia’. Lynch et al. (2000), Palmeri and Nosofsky (2001) and Burnett et al. (2005) found that ideals, or psychological extreme points, may define category goodness even in natural categories.<sup>1</sup> These studies also show that sometimes categorization can be based on ideals, and that people judge the ideal rather than the average members as the typical ones. Perhaps more interesting for this paper is the finding that when categories were learned in relation to alternative contrast categories, extreme members were counted as typical (cf. Ameels and Storms (2006)), and people were best able to categorize based on such ideals (cf. Goldstone et al. (2003)). This all suggests that if we want to model what it means to be a ‘real’, or typical,  $X$ , one should not just pick an average exemplar of type  $X$ .

If Rambo is not a prototypical man, he is certainly a *stereotypical* one. The Oxford English Dictionary defines a stereotype as a ‘widely held but fixed and oversimplified image or idea of a particular type of person or thing’. The so-called ‘social cognition approach’ to stereotypes (e.g. Schneider et al. (1979)), rooted in social psychology, views a social stereotype as a special case of a cognitive schema. Such schemas are intuitive generalizations that individuals routinely use in their everyday life, and entail savings on cognitive resources. Hilton and von Hippel (1996) define stereotypes as ‘mental representations of real differences between groups [...] allowing easier and more efficient processing of information. Stereotypes are selective, however, in that they are localized around group features that are the most distinctive, that provide the greatest differentiation between groups, and that show the least within-group variation.’ Thus, according to Hilton and von Hippel (1996), stereotypes are rather extreme representatives of a class.

Within social psychology, McCauley et al. (1980) have defined the following measure of how stereotypical  $x$  is for class  $X$ :  $\frac{P(x|X)}{P(x)}$ . An easy proof shows that this measure behaves monotone increasingly with respect to  $\log \frac{P(x|X)}{P(x|\neg X)}$ ,<sup>2</sup> meaning that the  $x$  with the highest value for the former notion also has the highest value for the latter notion. The latter notion goes back to Turing, and has been called the *weigh of evidence* of  $x$  for  $X$  by Good (1950). The same notion has been called the

<sup>1</sup>Of course, Plato already thought of universals as represented by *ideals* (the Forms).

<sup>2</sup>To show this, note first that  $P(x|X) - P(x)$  behaves monotone increasingly with  $P(x|X) - P(x|\neg X)$ .

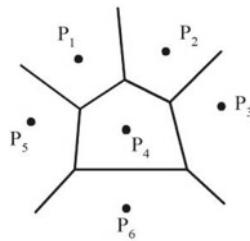
$$\begin{aligned} P(x|X) - P(x) &= P(x|X) - [(P(x|X) \times P(X)) + (P(x|\neg X) \times P(\neg X))] \\ &= P(x|X) - [\alpha P(x|X) + (1 - \alpha)P(x|\neg X)], \text{ with } 0 \leq \alpha \leq 1 \\ &= (1 - \alpha)P(x|X) - (1 - \alpha)P(x|\neg X) \\ &= \beta[P(x|X) - P(x|\neg X)], \text{ with } 0 \leq \beta \leq 1. \end{aligned}$$

Obviously,  $\frac{P(x|X)}{P(x)}$  behaves monotone increasingly with  $P(x|X) - P(x)$ , just as  $\frac{P(x|X)}{P(x|\neg X)}$  behaves monotone increasingly with  $P(x|X) - P(x|\neg X)$ . Given the nature of logarithmic functions, the latter, in turn, behaves monotone with  $\log \frac{P(x|X)}{P(x|\neg X)}$ .

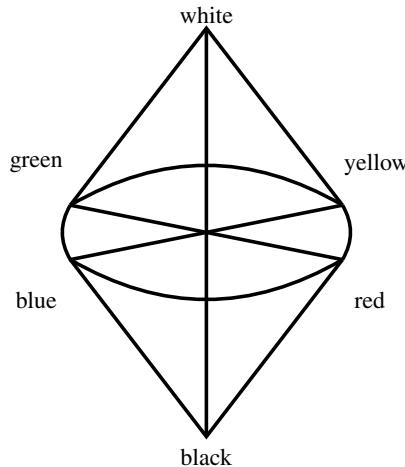
*representativeness* of  $x$  for  $X$  by Tenenbaum and Griffiths (2001). Adding things up, it all suggests that typical, or representative, members of their classes, are, in fact, their stereotypes, members that provide the greatest differentiation between the classes.

## 2 Typicality and Structured Meaning Spaces

Gärdenfors (2000) proposes that primitive categories (or natural properties) are always formed in contrast to alternative contrast categories in a priori given conceptual spaces. He suggests that—perhaps as a result—these basic categories are typically *convex sets*. A set  $X$  is convex if and only if for two arbitrary members  $x_1$  and  $x_2$  of  $X$ , any  $x_i$  that is somewhere between  $x_1$  and  $x_2$  is also a member of  $X$ . Gärdenfors claims that for primitive categories, the relevant conceptual spaces give rise to Voronoi tessellations. A Voronoi tessellation not only partitions a structured space into convex sets, it also has prototypes at the center of each convex set. Here is a typical example:

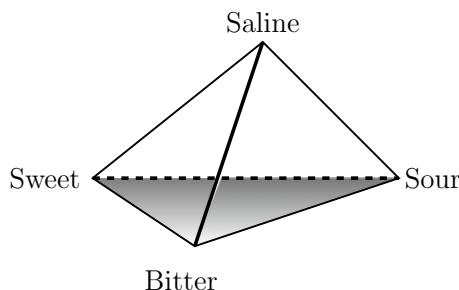


Two of the main examples discussed by Gärdenfors are colors and tastes. He claims that the color space and the phenomenological taste space give rise to Voronoi tessellations. We would like to question, however, whether the most typical colors and tastes are the central members, as proposed by Gärdenfors. First, consider one-dimensional spaces closed on at least one side. In linguistics (e.g. Kennedy and McNally (2005)), the meanings of contrastive adjective pairs such as ‘open’ and ‘closed’, ‘dry’ and ‘wet’ and ‘full’ and ‘empty’ are based on such one-dimensional spaces. The *endpoints* of such meaning spaces, however, will always be marked linguistically, by absolute adjectives, and thus be typical representatives of the classes such (absolute) adjectives denote. Second, inspection suggests that the focal points of the colors in the color space are not in the center. Below is a picture of the representations of colors as the full color spindle.



This picture strongly suggests two things: (i) that colors can be thought of as convex sets in the color space, and (ii) that the prototypes of the colors are (except for gray) always *at the edges* of the color spindle, and thus *not in the center* of the convex sets. Indeed, Regier et al. (2007) found that the best examples of English' white and black, respectively, are the lightest and darkest chips of a chart of colors.<sup>3</sup> Similarly, the so-called 'color emotion wheel' (from Sacharin et al. (2016), though not shown here), suggests that the color pixels which give rise to the highest emotions are on the edges of the color spindle (or circle, in this case). That picture also suggests that the pixels of the highest emotional value of the three basic colors red, blue and green, are *as far away from each other as possible*.

Finally, according to Henning (1916) the phenomenal gustatory space should be described by the following tetrahedron:




---

<sup>3</sup>A reviewer suggests that white and black are not real colors. This reviewer moreover suggests that 'true' colors only sit along the rim of the middle disc of the color spindle. All 'true' colors are maximally saturated, and only these should be considered. We are somewhat surprised about this suggestion. We agree that 'real', 'true', or stereotypical, red is red with full saturation, but we don't see any reason why we should limit the color space to full saturated colors to begin with. To us, this should be the *result* of an analysis, not the beginning.

Again, it seems that the basic tastes are convex regions of the relevant meaning space, and that their typical representatives are at the edge of the taste space, and far away from each other.

Bickerton (1981) already proposed that ‘simple’ expressions can only denote *connected*, or *convex*, regions of cognitive space, and hypothesized that the preference for convex properties is an *innate* property of our brains. Perhaps this is the case. Still, we would like to delve somewhat deeper and provide an analysis where convexity of meanings doesn’t have to be stipulated, but can be *explained*. The goal of this paper is to provide rational motivations for why standard meanings give rise to convex sets and why typical representatives are as far away from each other as possible.

Linguists like Jakobson (1941) and Martinet (1955) long observed that naturally occurring vowels in languages over the world are always far away from each other in the acoustic space available for vowels. Liljencrants and Lindblom (1972) showed that one can explain this linguistic ‘universal’ by adopting a principle of maximal perceptual contrast. Likewise, Regier et al. (2007) show that a model that categorizes the color space based on maximization of similarity within category and dissimilarity across categories gives rise to surprisingly accurate predictions for the predicted colors, and gives rise to categories as convex sets. Abbott et al. (2016) show that in trying to predict the focal colors, or the best examples of named color categories across many languages, a model making use of Tenenbaum and Griffiths’s (2001) notion of *representativeness* mentioned above outperforms several natural competitors such as models based on likelihood or on prototypes thought of as central members. Although very appealing, we feel that these explanations need to be based on the idea that language is used for communication between agents. This is the starting point of Lewis’s (1969) analysis of meaning making use of signaling games. In this paper we seek to motivate why meanings tend to be convex and why extreme exemplars of these meanings, or categories, are considered to be representative by making use of such signaling games.

Jäger (2007) and Jäger and van Rooij (2007) introduced so-called sim-max games, signaling games using an Euclidean meaning space with a similarity-based utility function. They show that by using a simple learning dynamic the evolved equilibria of these games give rise to descriptive meanings which are convex sets.<sup>4</sup> For *sim-max games*, it is shown as well that with uniformly distributed points in the meaning spaces, the imperative meanings derived from the equilibria will be in the center of their descriptive meanings, and can be thought of as prototypes. As argued above, we indeed want an explanation of convex meanings, but now with typical representatives as extremes.<sup>5</sup> Zuidema and de Boer (2009) observed that Liljencrants and Lindblom (1972)’s explanation of naturally occurring vowels as extremes in the acoustic space in terms of maximal contrast makes game theoretical sense in a noisy environment. In this paper we would like to provide a game theoretical

<sup>4</sup> Elliot Wagner (p.c.) has shown, however, that this does not hold in general, if a more standard evolutionary dynamic is used.

<sup>5</sup> One might think that the problem can be solved by adopting a non-flat probability distribution. As observed by Franke (2012), however, this won’t do.

explanation of a phenomenon involving maximal contrast as well. But there is an important difference: whereas in phonology the contrast involves the *signals*, in our case the contrast involves the *meanings* of the signals. For simple one-dimensional meaning spaces, Lipman (2009) already provided such a game theoretical derivation, not making use of similarity or confusability at all. Surprisingly enough, his analysis even explains convexity. Unfortunately, we don't see how to extend his derivation to more complex spaces. Franke (2012) does explain the preference for extreme points in multi-dimensional spaces.<sup>6</sup> However, he does so by doing it, so to say, in terms of a derived preference for extremes in one-dimensional spaces. What we would like to do is, we think, more ambitious: to explain the preference for the extremes in one go. We think that something like this is required to provide a natural explanation of the preference for extremes in complex spaces where the dimensions are not obviously made up of previously given dimensions that are independent of each other. Such a dependence of the dimensions we find, for instance, in the color space which Gärdenfors (2000) takes to be consisting of a set of *integral* dimensions.

### 3 Extremes and Iterated Best Response

One way to understand why languages exhibit the properties they do is by analyzing them in the context of cooperative social reasoning. That is, by taking the idea seriously that language is used for communication between interlocutors, and that these interlocutors will reason about each other's linguistic choices to reach mutual understanding (e.g., Lewis, 1969; Grice 1975; Parikh 1991; Rooy van 2004; Benz et al. 2005). To illustrate how such a process of mutual reasoning may naturally lead to convex meanings with extreme typical representatives, this section sketches out the predictions of the Iterated Best Response (IBR) model (Franke 2009; Franke and Jäger 2014) on these matters.

At its core, IBR aims to explain linguistic outcomes in a Gricean fashion: as outcomes of mutual reasoning about rational language use. Formally, patterns of language use can be represented by mappings from messages (utterances) to states (meanings) in the case of receivers,  $\rho: M \rightarrow T$ ; and by mappings from states to messages for senders,  $\sigma: T \rightarrow M$ . Plainly put, these are comprehension and production strategies that tell us how two interlocutors behave. That sender and receiver are rational means that, given (their beliefs about) another interlocutor's behavior, they will try to maximize communicative success. If, e.g., the sender believes the chances of the receiver interpreting utterance  $m_1$  correctly to be higher than those of utterance  $m_2$ , she will send the former. Letting  $R$  and  $S$  be the set of all receiver and sender strategies, the set of best responses to a sender/receiver belief is defined as follows:

---

<sup>6</sup>Explaining convexity is not aimed for in Franke (2012).

$$\text{BR}(\sigma_b) = \{\rho \in R \mid \forall m : \rho(m) \in \text{argmax}_{t \in T} EU_R(t, m, \sigma_b)\}; \quad (1)$$

$$\text{BR}(\rho_b) = \{\sigma \in S \mid \forall t : \sigma(t) \in \text{argmax}_{m \in M} EU_S(t, m, \rho_b)\}, \quad (2)$$

where  $\sigma_b$  and  $\rho_b$  are the receiver's, respectively the sender's, beliefs about her interlocutor's behavior and  $EU(t, m, \cdot)$  codifies the expected utility of either interpreting a message  $m$  as  $t$  or sending a message  $m$  in state  $t$  (see below).

Equations (1) and (2) may look unwieldy at first glance, so let us unravel them before moving on. A belief about a sender/receiver strategy is an expectation of how this sender/receiver will act given a state/message. Beyond the fact that they are beliefs about another agent's behavior, these are just mappings from states/messages to messages/states as well. A best response to an interlocutor's (expected) behavior is the strategy that will ensure the best payoff from an interaction with such an interlocutor: the one with the highest expected utility. There might be many ways to use language that maximize utility conditional on a particular belief  $\sigma_b$  or  $\rho_b$ ; the sets  $\text{BR}(\sigma_b)$  and  $\text{BR}(\rho_b)$  collect them all.

Having identified the set of best courses of action given a belief about an interlocutor's behavior, we still need to distill from them how an agent should act. For convenience, we write the resulting strategies as behavioral ones. In words, a sender's strategy  $\sigma$  is the one that sends a message  $m$  in state  $t$  if there is a best response  $\sigma' \in \text{BR}(\rho_b)$  that sends it. Otherwise, message  $m$  is not sent in  $t$ . Formally,  $\sigma(m \mid t, \rho_b) = 1 / |\{m' \mid \sigma'(m') = t; \wedge \sigma' \in \text{BR}(\rho_b)\}|$  if there is a strategy  $\sigma' \in \text{BR}(\rho_b)$  such that  $\sigma'(t) = m$ , and otherwise 0. Analogously for  $\rho(t|m, \sigma_b)$ , with the additional proviso that if a message is not believed to be sent at all, the receiver will pick an interpretation at random (cf. Franke and Jäger 2014).

As a final ingredient, we need to specify what sender and receiver care about. Assuming that interlocutors have no preferences over messages and that all they care about is faithful information transfer, utility can be captured by a single function that tracks how closely sender state and receiver interpretation match; e.g.,  $\delta(t, t') = 1$  iff  $t = t'$  and otherwise 0. We then have

$$EU_R(t, m, \sigma_b) = \sum_{t'} \frac{Pr(t')\sigma_b(m|t')}{\sum_{t''} Pr(t'')\sigma_b(m|t'')} \delta(t', t); \quad (3)$$

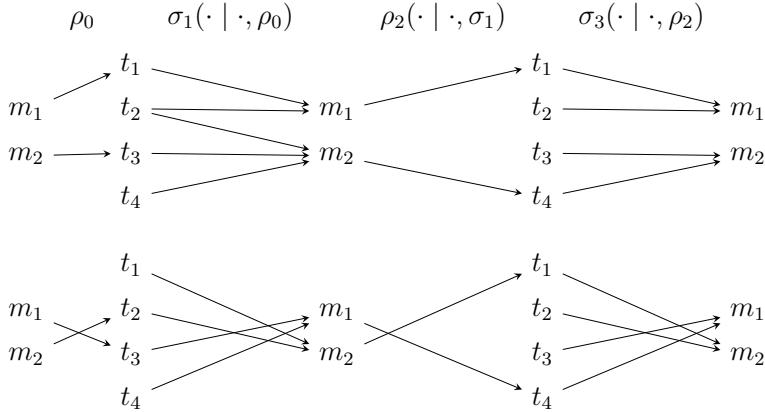
$$EU_S(t, m, \rho_b) = \sum_{t'} \rho_b(t'|m) \delta(t, t'). \quad (4)$$

In words, the expected utility of sending/interpreting message  $m$  given state  $t$  is just the average of our communicative success given our beliefs about our interlocutor's linguistic behavior. That is to say, expected utility gives us the average payoff we expect when producing or comprehending, conditional on our beliefs about our communicative partner. As stated in (1) and (2), best responses are made up of those strategies that maximize expected utility; those that guarantee the best outcome based on what we care about.

All of this is just to formally capture the idea that a message is sent only in states in which it is believed to have the highest chances to be understood; and that, analogously, a receiver interprets a message as the state that she believes is most likely to be conveyed by it. If there are many optimal choices, players pick randomly from them. If a choice has to be made but none is optimal they pick at random from the entire pool of actions at their disposition. From here, we just need to consider the consequences of nesting beliefs to arrive at pragmatic reasoning: reasoning about the reasoning (and so on) of others to inform our linguistic choices. Formally, a level- $n + 1$  reasoner in IBR is defined as acting upon the belief that her interlocutor is of level- $n$  with reasoning levels starting at  $n = 0$ . Put differently, we have that  $\sigma_{n+1}(\cdot | \cdot, \rho_n)$  and  $\rho_{n+1}(\cdot | \cdot, \sigma_n)$ .

Beliefs about an interlocutor's strategy at level 0 are usually constrained or biased in some way to start the reasoning chain. If just any belief were permitted, meaningful inference would seldom get off the ground (cf. Sect. 1.2 Franke 2009). Let us consider a simple case in which the sender has seen how the receiver interprets messages and the receiver is aware of this. For instance, she has seen the receiver interpret the utterance *tall woman* as an entity of a particular height and *small man* as an entity of another height. As we shall see, we need not constrain this receiver strategy beyond requiring that it associates each message with a distinct information state. Mutual awareness of this arbitrary separating strategy suffices to lead to the adoption of convex strategies with extreme typical representatives as long as extremes are salient. Saliency could be cashed out in different ways: It may be that extremes are focal points that draw the attention of reasoners due to their psychological noteworthiness relative to other states ( cf. Schelling 1980; Mehta et al. 1994); or it might be that extremes confer a functional advantage and attract the reasoners by virtue of their drive to maximize expected utility. The latter might happen, e.g., if perception is noisy in that states that are near to each other are easily confused. This would make extremes attractive in virtue of their special position at the edge of a space, making them the least confusable (see, e.g., Franke et al. 2011, Gibson et al. 2013, Franke and Correia 2018 for other proposals where noise, or error, has been argued to play an explanatory role). Abstracting away from the details of particular noise signatures, their consequences can be captured by a graded utility function that is inversely proportional to a distance measure over the state space under the assumption that coordinating on extremes confers a higher utility than coordinating on less extreme points. We background the details of this function because these two general conditions are sufficient to illustrate our argument. In which way extreme points are salient is ultimately an empirical issue. At this stage proposing a particular function seems too strong a commitment in light of these unknowns.

With these notions at hand, consider the case of four heights,  $T = \{1, 2, 3, 4\}$ , and two messages,  $M = \{m_1, m_2\}$ . Figure 1 illustrates how mutual reasoning can lead to convex strategies with extreme typical representatives when reasoning over two initial receiver strategies  $\rho_0$ . Intuitively, a level-1 rational sender strategy against a belief of her interlocutor's behavior,  $\sigma_1(\cdot | \cdot, \rho_0)$ , will first ensure that messages sent in a state correspond to correctly interpreted messages;  $t_1 \mapsto m_1$  and  $t_3 \mapsto m_2$  in the upper example of Fig. 1; and  $t_2 \mapsto m_2$  and  $t_3 \mapsto m_1$  in the lower one. Second, remaining

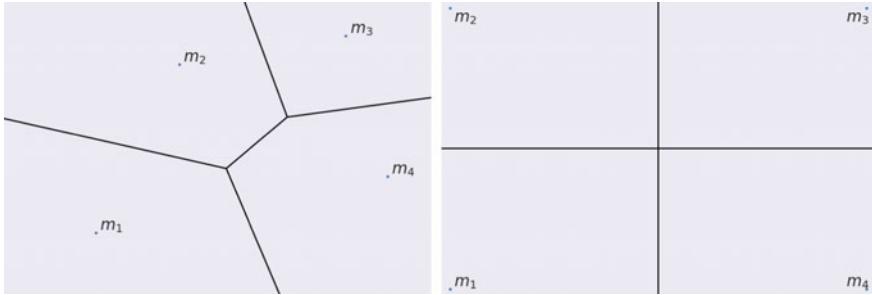


**Fig. 1** Illustration of IBR-sequence for two separating initial receiver strategies  $\rho_0$ . Depicted outcomes correspond to endpoints of the reasoning process

states will be associated with messages whose interpretation is closest to them. In the upper example in Fig. 1 state  $t_2$  lies in between  $\rho_0$ 's interpretation of  $m_1$  and  $m_2$ , so it is associated with both. A (level-2) receiver who reasons about such a message allocation will naturally associate her messages with the interpretations that are most rewarding: the extremes. Subsequent sender reasoning leads to the association of remaining states such that the state space is partitioned into convex regions. As noted above, this may, e.g., be a consequence of reasoned noisy perception or that of a particular graded utility function. More iterations will not change the sender and receiver strategies anymore. They are in equilibrium.

Just as in Lewis, (1969), we can ascribe two types of meanings to a message in these equilibrium pairs: its *descriptive meaning* is the set of states in which this message is sent and its *imperative meaning* is the response to this message by the receiver. Just as in standard sim-max games, descriptive meanings are now convex sets. But whereas imperative meanings in Jäger (2007) and Jäger and van Rooij (2007) were central points, i.e., prototypes, now they are extreme points, i.e., stereotypes.

This outcome is not limited to one-dimensional spaces such as this ordering of heights. Instead, it obtains in any discrete space with a distance measure, should there be at least as many extreme points as messages. For instance, the color spindle, the taste space, or any discrete subset of a multi-dimensional interval. In any such space, mutual reasoning will iteratively lead to a rational receiver's association of (at least some) messages with extremes. A rational sender follows suit by uniquely identifying extremes with these messages, as well as by improving the space's tessellation with respect to these associations. This process continues as long as the receiver has not yet associated each message with an extreme, being driven by the improved partition each round of back-and-forth reasoning provides. In the end, mutual reasoning bottoms out with convex sender strategies with extreme typical representatives



**Fig. 2** Illustration of IBR-sequence in a two-dimensional space. Labeled nodes in the left-hand picture depict an initial receiver strategy  $\rho_0$ . The resulting convex sender strategy  $\sigma_1(\cdot | \cdot, \rho_0)$  corresponds to the four regions enclosing each node. Labeled nodes in the right-hand picture correspond to  $\rho_2$  and regions enclosing them depict  $\sigma_3$

for receiver strategies. Figure 2 sketches out how convex descriptive meanings and extreme imperative ones result from mutual reasoning in such a space.

In the previous section we mentioned that best examples of named color categories are well-predicted by a model based on the following measure of representativeness,  $\log \frac{P(x|X)}{P(x|\neg X)}$ , which is very similar to a measure used to define stereotypicality. It is interesting to observe that our game-theoretical analysis predicts that the imperative meaning of messages in equilibrium are the most representative ones for their descriptive meanings. To show this, one has to think of  $P(t|m)$  and  $P(t|\neg m)$  either in terms of sender strategies or in terms of receiver strategies. In the former case, one can interpret  $P(t|m)$ , for instance, as the probability that  $t$  is the actual state if  $m$  is sent. However, it is easier to think of  $P(t|m)$  and  $P(t|\neg m)$  in terms of receiver strategies. In that case,  $P(t|m)$ , for instance, is just  $\rho(t|m, \sigma_b)$ , with  $\rho$  and  $\sigma$  as the equilibrium receiver and sender strategies, respectively. Once one assumes that senders and receivers use a *quantal* instead of a *maximizing* best response function,<sup>7</sup> in the upper example of Fig. 1, for instance,  $t_1$  and  $t_4$  maximize  $\log \frac{\rho(t|m_1, \sigma_b)}{\rho(t|\neg m_1, \sigma_b)}$  and  $\log \frac{\rho(t|m_2, \sigma_b)}{\rho(t|\neg m_2, \sigma_b)}$ , respectively, and are thus predicted to be the most representative states for  $m_1$  and  $m_2$ . In other words, they are the *stereotypes* of the (descriptive) meanings of the

<sup>7</sup>The need for quantal best response is due to a technical complication resulting from the use of maximizing expected utility: it often causes the measure of representativeness to be undefined. To see this, notice that the most representative, or stereotypical, state for message  $m$  would now be  $\text{argmax}_{t \in T} \log \frac{\rho(t|m, \sigma_b)}{\rho(t|\neg m, \sigma_b)}$ . But as illustrated in, for instance, the upper example of Fig. 1,  $\rho_2(t_1|m_2, \sigma_1) = 0$ , meaning that the denominator of  $\frac{\rho_2(t_1|m_1, \sigma_1)}{\rho_2(t_1|\neg m_1, \sigma_1)}$  is 0, which makes the fraction undefined. This problem is solved if we make sure that for no  $t$  and  $m$  it ever will be the case that  $\rho(t|m, \sigma) = 0$ . This is what comes out if we assume that instead of being expected utility maximizers, senders and receivers choose probabilistically modeled by quantal response functions (QRFs). These functions are motivated by the idea that (perhaps due to observation errors) decision makers sometimes make mistakes in choosing their best action. These functions are popular in behavioral economics and are gaining popularity in linguistics as well, as they more readily connect rational language use models with empirical data (see, e.g., Franke et al. 2011; Frank and Goodman 2012; Franke and Jäger 2016).

messages. This result is not limited to our simple example using a one-dimensional meaning space, but generalizes to more-dimensional spaces: *stereotypes follow from (boundedly) rational language use.*

## 4 Conclusion and Outlook

In this paper we followed Gärdenfors and others in the assumption that (simple) properties denote convex sets in conceptual spaces, but argued that typical representatives of categories are (many times) extreme rather than central members of such categories, i.e., stereotypes. Moreover, we provided a rational motivation for convexity of meaning and of stereotypes as typical representatives making use of game theory.

We believe that these motivations are interesting for more general linguistic reasons. For instance, it is not uncommon to believe that generic sentences like ‘Birds fly’ and ‘Sharks are dangerous’ express typicalities and it is well-known that generics are excellent tools to express and generate stereotypes. In Rooij van and Schulz (2020) an analysis of generic sentences is proposed based on *contingency*, a measure of representativeness adopted from causal associative learning theory that behaves monotone increasingly with the measures of stereotypicality and representativeness discussed in this paper. This suggests that we could provide a game theoretical motivation for generic language use as well. There is at least one complication, though. Whereas we thought of stereotypes as *members* of a category, for Rooij van and Schulz (2020) it is crucial to think of stereotypes as *sets of* perhaps mutually inconsistent *features*. In the future we would like to see how crucial this distinction is.

In this paper—just as in Jäger (2007) and others—we have fixed the number of messages that play a role in the game beforehand, which determined the number of cells in the resulting partition of the meaning space in equilibrium. Intuitively, that should not be the case: in how many cells the meaning space will be partitioned should be an *outcome* of the game as well, depending on the structure of the meaning space and the utility of each partition. Corter and Gluck (1992) defined a notion of *category utility* to derive Rosch’s so-called ‘basic-level’ categories. It is interesting to observe that this notion is closely related to the notions of ‘representativeness’, ‘contingency’ and ‘stereotypicality’ discussed above. In the future we would like to explain natural partitions of different types of meaning spaces, making use of this notion of *category utility*.

## References

- Abbott, J., Regier, T., & Griffiths, T. L. (2016). Focal colors across languages are representative members of color categories. *Proceedings of the National Academy of Sciences*, 113, 11178–11183.

- Ameels, E., & Storms, G. (2006). From prototypes to caricatures: Geometrical models for concept typicality. *Journal of Memory and Language*, 55, 402–421.
- Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 629–654.
- Benz, A., Jäger, G., & van Rooij, R., (Eds.) (2005). *Game theory and pragmatics*. Springer.
- Bickerton, D. (1981). *Roots of language*. Karoma Publishers.
- Burnett, R., Medin, D., Ross, N., & Blok, S. (2005). Ideal is typical. *Canadian Journal of Experimental Psychology*, 59(3–10).
- Corter, J., & Gluck, M. (1992). Explaining basic categories: Feature predictability and information. *Psychological Bulletin*, 111, 291–303.
- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084), 998.
- Franke, M. (2009). *Signal to act: Game theoretic pragmatics*. Ph.D. thesis, University of Amsterdam.
- Franke, M. (2012). On scales, salience and referential language use. In Aloni, M., Roelofsen, F., & Schulz, K., (Eds.), *Proceedings of Amsterdam Colloquium 2011* (pp. 311–320). Springer.
- Franke, M., & Correia, J. (to appear). Vagueness and imprecise imitation in signalling games. *British Journal for the Philosophy of Science*.
- Franke, M., & Correia, J. (2018). Vagueness and imprecise imitation in signalling games, *British Journal for the Philosophy of Science*, 69(4), 1037–1067
- Franke, M., & Jäger, G. (2014). Pragmatic back-and-forth reasoning. In *Pragmatics, Semantics and the Case of Scalar Implicatures*: Nature Publishing Group.
- Franke, M., & Jäger, G. (2016). Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft*, 35(1).
- Franke, M., Jäger, G., & van Rooij, R. (2011). Vagueness, signaling and bounded rationality. In B. Onoda, & D. J. McReady (Eds.), *New frontiers in artificial intelligence* (pp. 45–59). Springer.
- Gärdenfors, P. (2000). *Conceptual spaces*. The Geometry of Thought: MIT Press.
- Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, 110(20), 8051–8056.
- Goldstone, R., Steyvers, M., & Rogosky, B. (2003). Conceptual interrelatedness and caricatures. *Memory and Cognition*, 31, 169–180.
- Good, I. (1950). *Probability and the weighing of evidence*. London: Griffin.
- Grice, P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). Nature Publishing Group.
- Hampton, J. (1981). An investigation of the nature of abstract concepts. *Memory and Cognition*, 9, 149–156.
- Henning, H. (1916). Die Qualitätenreihe des Geschmacks. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 74, 203–219.
- Hilton, J., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, 47, 237–271.
- Jäger, G. (2007). The evolution of convex categories. *Linguistics and Philosophy*, 30, 551–564.
- Jäger, G., & van Rooij, R. (2007). Language structure: Psychological and social constraints. *Synthese*, 159, 99–130.
- Jakobson, R. (1941). *KinderSprache, aphasie und allgemeine lautgesetze*. Uppsala.
- Kennedy, C., & McNally, L. (2005). Scale structure and the semantic typology of gradable predicates. *Language*, 81, 1–37.
- Lewis, D. (1969). *Convention: A philosophical study*. Cambridge: Harvard University Press.
- Liljencrants, J., & Lindblom, B. (1972). Numerical simulations of vowel quality systems: The role of perceptual contrast. *Language*, 48, 839–862.
- Lipman, M. (2009). Why is language vague? Unpublished manuscript.
- Lynch, E., Coley, J., & Medin, D. (2000). Tall is typical: Central tendency, ideal dimensions, and graded category structure among tree experts and novices. *Memory and Cognition*, 28, 41–50.
- Martinet, A. (1955). *Économie des changements phonétiques*. Berne: Francke.

- McCauley, C., Stitt, C., & Segal, M. (1980). Stereotyping: From prejudice to prediction. *Psychological Bulletin, 87*, 195–208.
- Mehta, J., Starmer, C., & Sugden, R. (1994). Focal points in pure coordination games: An experimental investigation. *Theory and Decision, 36*(2), 163–185.
- Palmeri, T., & Nosofsky, R. (2001). Central tendencies, extreme points, and prototype enhancement effect in ill-defined perceptual categorization. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 54*, 197–235.
- Parikh, P. (1991). Communication and strategic inference. *Linguistics and Philosophy, 14*(5), 473–514.
- Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences, 104*, 1436–1441.
- Rooij, van R. & Schulz, K. (2020). Generics and typicality: A bounded rationality approach. *Linguistics and Philosophy, 43*(1), 83–117.
- van Rooy, R. (2004). Signaling games select horn strategies. *Linguistics and Philosophy, 27*, 493–527.
- Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In *Cognitive development and the acquisition of language*. Academic Press.
- Sacharin, V., Schlegel, K., & Scherer, K. R. (2016). Geneva emotion wheel rating study. *Report*.
- Schelling, T. C. (1980). *The strategy of conflict*. Harvard University Press.
- Schneider, D., Hastorf, A., & Ellsworth, P. (1979). *Person Perception*. Adison-Wesley Pub. Co.
- Tennenbaum, J., & Griffiths, T. (2001). The rational basis of representativeness. In J. Moore, & K. Stenning, (Eds.), *Proceedings of the 23th Annual Conference of the Cognitive Science Society*, pp. 1036–1041.
- Zuidema, W., & de Boer, B. (2009). The evolution of combinatorial phonology. *Journal of Phonetics, 37*, 125–144.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Grading Similarity



Carla Umbach and Helmar Gust

**Abstract** There are numerous words across languages expressing similarity or indistinguishability. In this paper, three types of similarity expressions in German and English are compared—*ähnlich/similar*, *so/such*, and *gleich/same*. They differ in a number of respects, one of them being gradability: While *ähnlich/similar* are gradable, *so/such* as well as *gleich/same* are not. The analysis in this paper starts from the analysis of German *so* as a demonstrative expressing similarity (instead of identity) to its demonstration target (Umbach and Gust 2014). It is suggested that the meaning of the three types of similarity expressions is based on a common similarity relation, while differences in meaning are provided by constraints referring to the selection of dimensions of comparison and preconditions of usage. The focus of the paper is on gradability and on the question of what it means for a pair of items to be more similar than another pair. An analysis in the spirit of Klein (1980) is presented accounting for the fact that *ähnlich/similar* are gradable while neither *so/such* nor *gleich/same* are. The formal framework makes use of representations based on attribute spaces and classifiers, where representations may be of different granularity.

**Keywords** Similarity · Sameness · Dimensions of comparison · Direct reference · Gradability

## 1 Introduction

There are numerous words across languages expressing that items are similar or indistinguishable in some sense, for example in German and English *ähnlich/similar*, *so/such*, and *gleich/same*. It seems reasonable to assume that the common core of the meaning of these words is a relation of similarity, which is considered in Cognitive

---

C. Umbach (✉)

Leibniz-Zentrum Allgemeine Sprachwissenschaft Berlin/Universität Köln, Berlin, Germany

e-mail: [carla.umbach@uni-koeln.de](mailto:carla.umbach@uni-koeln.de)

H. Gust

Institut Für Kognitionswissenschaft, Universität Osnabrück, Osnabrück, Germany

e-mail: [hgust@uos.de](mailto:hgust@uos.de)

© The Author(s) 2021

S. Löbner et al. (eds.), *Concepts, Frames and Cascades in Semantics, Cognition and Ontology*, Language, Cognition, and Mind 7,  
[https://doi.org/10.1007/978-3-030-50200-3\\_17](https://doi.org/10.1007/978-3-030-50200-3_17)

Science as “... an organizing principle by which individuals classify objects, form concepts, and make generalizations” (Tversky 1977, p. 327). Still, there are significant differences between similarity expressions, one of them being gradability: While *ähnlich* and *similar* are gradable, *so* and *such* as well as *gleich* and *same* are not, see (1)–(3).<sup>1</sup>

- (1) (speaker points to Berta’s haircut):
  - a. Anna hat auch so einen Haarschnitt / \*... mehr so einen Haarschnitt als Claire.
  - b. Anna has such a haircut, too. / \*... more such a haircut than Claire has.
- (2) a. Annas Haarschnitt ist dem von Berta ähnlich. / ... dem von Berta ähnlicher als der von Claire.
  - b. Anna’s haircut is similar to Berta’s haircut. / ... more similar to Berta’s haircut than Claire’s haircut is.
- (3) a. Annas Haarschnitt ist der gleiche wie Bertas./ \*... mehr der gleiche wie Bertas als der von Claire.
  - b. Anna’s haircut is the same as Berta’s haircut. / \*... more the same as Berta’s haircut than Claire’s haircut is.

The starting point of this paper is the analysis of the German demonstrative *so* in Umbach and Gust (2014) arguing that German *so* as well as, e.g., Polish *tak* and English *such* are *similarity demonstratives*, that is, demonstratives expressing similarity (instead of identity) to the target of the demonstration (see Sect. 2). The similarity analysis is spelled out with the help of multi-dimensional attribute spaces defining similarity as indistinguishability with respect to, basically, a set of dimensions of comparison.

German *ähnlich* and English *similar* express similarity, too. But while *so* and *such* are demonstratives, *ähnlich* and *similar* are two-place predicates, and while similarity as denoted by *so* and *such* is reflexive,<sup>2</sup> it will be shown that this is not the case for *ähnlich* and *similar*. The most challenging difference, however, is gradability, which will be in focus in this paper.

Considering their scale structure, *ähnlich* and *similar* are clearly not open scale—increase of similarity is not open-ended. But at the same time they resist common tests for being upper-closed (see Kennedy and McNally 2005). For example, combination with *vollkommen/completely* yields heavily marked results. Intuitively, however, there is a maximum for *ähnlich* and *similar* which is expressed by the adjectives *gleich* and *same*, see (4).

---

<sup>1</sup>There is *mehr so* ('more so') in the sense of *eher so* ('rather so') which is, however, a hedging construction instead of a comparative, as is evident from the fact that the standard parameter is *wie* instead of *als*: *Anna hat mehr/eher so einen Haarschnitt wie Claire/\*als Claire* ('Anna has more such a haircut as/than Claire.').

<sup>2</sup>For similarity expressed by *so* and *such* it holds that  $\forall x \in D. SIM(x,x)$ .

- (4) a. Anna hat ??? einen vollkommen ähnlichen / <sup>ok</sup> den gleichen Haarschnitt wie Berta.  
 b. Anna has ??? a completely similar haircut to Berta / <sup>ok</sup> the same haircut as Berta.

In this paper, we start from the idea that the meaning of the three types of similarity expressions—*so/such*, *ähnlich/similar*, and *gleich/same*—is based on a single similarity relation. Differences in meaning are characterized in terms of additional constraints. The research questions addressed in this paper will be

- (i) What are the distinctive characteristics of the three types of similarity expressions?
- (ii) What kind of gradable adjectives are *ähnlich* and *similar*? How to explain that *ähnlich/similar* are gradable while neither *so/such* nor *gleich/same* are?
- (iii) What does it mean for two items *a* and *b* to be more similar than *c* and *d*? How to implement the gradability of *ähnlich/similar*?

In this paper, we will consider only nominal phrases (ignoring e.g., *ähnlich aussehen/look similar* and also *ähneln/resemble*; for *resemble* see Meier 2009) and we will only consider anaphoric/deictic uses (ignoring reciprocal constructions like *Anna and Berta are similar*, see footnote 11 in Sect. 3). Since the German and English expressions under consideration are close in meaning and distribution they will be analyzed in parallel.

This paper is organized as follows: In Sect. 2, the similarity analysis for *so/such* will be outlined as far as required in the subsequent sections. In Sect. 3, differences in distribution and meaning between the three types of similarity expressions will be explored. In Sect. 4, an analysis will be suggested accounting for the gradability of *ähnlich/similar* which is inspired by Klein (1980). Formal details are provided in the Appendix.

## 2 Similarity Demonstratives

There is a class of demonstratives found across languages modifying verbal, nominal and/or degree expressions, for example German *so/solch*, English *such*, Polish *tak* and Turkish *böyle* (see König and Umbach 2018). Some of them are uniform across categories, like German *so* and Polish *tak*; others are restricted to particular syntactic categories, like English *such*. In (5), German *so* and English *such* modify a noun.

- (5) (speaker points to a table):  
 a. So einen Tisch hat Anna auch.  
 b. Anna has such a table, too.

In Umbach and Gust (2014), demonstratives like *so/such* are called *similarity demonstratives* and are analyzed in a framework spelling out similarity as indistinguishability with respect to given dimensions of comparison. This section provides a summary of the analysis and a brief overview of the formal framework. Details are provided in the Appendix.

The analysis starts from the common idea that the target of the demonstration is an individual or event. But while standard demonstratives like *this* denote identity between the demonstration target and the referent (as is in-built in Kaplan's 1989 system), similarity demonstratives denote similarity rather than identity. Accordingly, *so/such* include a deictic component and a similarity component which jointly create sets of items similar to the target of demonstration. For example, *so ein Tisch/such a table* in (5) denote a set of tables similar to the table pointed at. This analysis entails that *so/such* are directly referential in the sense of Kaplan, which will be one key point in distinguishing *so/such* from *ähnlich/similar* and *gleich/same* in Sect. 3.

Similarity depends on dimensions of comparison.<sup>3</sup> The selection of the relevant dimensions is another key point in comparing the three varieties of similarity expressions. In the formal framework (Gust and Umbach 2015, Gust and Umbach to appear), dimensions of comparison define multidimensional attribute spaces and are equipped with measure functions mapping individuals to points in those spaces. Dimensions and measure functions are two components of what is called a *representation*. The third component is a set of *classifiers*, which are predicates on points in attribute spaces. They can be seen as defining a “grid”<sup>4</sup> where points within a cell are indistinguishable. Classifiers derived from basic ones by logical operations provide coarser (by disjunction) or finer granularity (by conjunction), which will be essential in devising a gradable notion of similarity in Sect. 4.2. Slightly simplifying, a representation  $\mathcal{F}$  is defined as a quadruple including a domain  $D$ , an attribute space  $F$ , a measure function  $\mu: D \rightarrow F$  and a set of classifiers  $P^*$ ,  $\mathcal{F} = \langle F, \mu, P^*, D \rangle$  (see Appendix, Definition 2).

Similarity is defined as a three-place relation combining two individuals to be compared and a representation,  $SIM(x, y, \mathcal{F})$ , such that two individuals are similar relative to a representation if and only if the points in the attribute space they are mapped to are indistinguishable relative to the given set of classifiers (Appendix, Definitions 3 and 4). Similarity defined in this way is an equivalence relation.<sup>5</sup>

Consider, for example, the phrases *so einen Tisch/such a table* in (5). The semantic interpretation is shown in (6). Let us assume, for the sake of the example, that relevant dimensions of comparison are *HEIGHT*, *MATERIAL*, *LEGS*, and *EXTRAS*, and that tables are “measured” by the function in (7). Now suppose that the table the speaker points

<sup>3</sup>Without taking dimensions of comparison into account, similarity runs the risk of being trivial, which is nicely demonstrated in Goodman (1972).

<sup>4</sup>The term “grid” is not to be misunderstood as implying a distance-based notion of similarity.

<sup>5</sup>Counter-arguments (going back to Tversky 1977) against defining similarity as an equivalence relation cannot in general be maintained, see footnote 23 in Sect. 4.2.

at is mapped to  $\langle 55\text{ cm}, \text{metal}, 4, \{\} \rangle$  and the set of classifiers is such that points within a range of *HEIGHT*:40–60; *MATERIAL*:{*metal*, *plastics*}; *LEGS*:2–4; *EXTRAS*:{} cannot be distinguished. Then (5) is true iff Anna’s table is mapped to a point within this range.<sup>6,7,8</sup>

$$(6) [[\text{so ein Tisch / such a table}]] = \lambda x. \text{table}(x) \& \text{SIM}(x, t, \mathcal{F})$$

$$(7) \mu_{\text{table}} : D \rightarrow \text{HEIGHT} \times \text{MATERIAL} \times \text{LEGS} \times \text{EXTRAS}$$

$$\text{where } \mu_{\text{table}}(x) = \langle \mu_{\text{height}}(x), \mu_{\text{material}}(x), \mu_{\text{legs}}(x), \mu_{\text{extras}}(x) \rangle$$

$$\text{and } \text{HEIGHT: } \mathbb{R}^+$$

$$\text{MATERIAL: } \{\text{wood, metal, plastics, ...}\}$$

$$\text{LEGS: } \{1 \dots 10\}$$

$$\text{EXTRAS: } \wp\{\text{extendable, height-adjustable, coating, ...}\}$$

According to the similarity analysis, demonstratives like *so/such* create classes of similar items, e.g. similar tables. There is some evidence that in the nominal and verbal case (though not in the adjectival case) these similarity classes constitute ad-hoc kinds (see Umbach and Gust 2014). Anderson and Morzycki (2015) present an alternative analysis claiming that demonstratives like German *so*, English *such* and Polish *tak* are pro-kind expressions, adapting Carlson’s (1980) kind-referring analysis of *such*. The final results of the two accounts are fairly close (in the case of nominal and verbal phrases). However, Umbach and Gust not just postulate that there are kinds denoted by *so* phrases, but in addition show how these kinds emerge, namely by similarity. Moreover, by referring to a common similarity relation, this framework offers a basis to compare different types of similarity expressions, which is the topic in this paper.

Finally, it is important to note that the notion of similarity in this framework is qualitative (property-based), unlike that in Gärdenfors’ (2000) conceptual spaces which

<sup>6</sup>Note that this approach does not classify objects as tables but instead creates subsets of similar tables.

<sup>7</sup>Regarding ex. (6), there are two options to interpret adnominal *so/such*: Either *so/such* are considered as modifiers of the indefinite determiner, or they are considered as modifiers of the nominal (and have been moved into the prenominal position). The first option yields the interpretation in (a) and the second the one in (b). Since the resulting quantifiers are identical and in German the prenominal position is licensed for *solch* (‘such’)—*ein solcher Tisch* ‘a such table’—we will analyze *so/such* in this paper as nominal modifiers, as in (b) and (6). This option facilitates comparison with *ähnlich/similar* and *gleich/same* because they occur as nominal modifiers, too.

$$a. [[\text{so/such ein/a}]] = \lambda P. \lambda Q. \exists x. \text{SIM}(x, t, \mathcal{F}) \& P(x) \& Q(x)$$

$$[[\text{so/such ein/a}]] ([[Tisch/table]]) = \lambda Q. \exists x. \text{SIM}(x, t, \mathcal{F}) \& \text{table}(x) \& Q(x)$$

$$b. [[\text{so/such Tisch/table}]] = \lambda x. \text{SIM}(x, t, \mathcal{F}) \& \text{table}(x)$$

$$[[\text{ein/a}]] ([[so/such Tisch/table]]) = \lambda Q. \exists x. \text{SIM}(x, t, \mathcal{F}) \& \text{table}(x) \& Q(x)$$

<sup>8</sup>On a related note, in ex. (6), German *so*, but not English *such*, may modify verbal and adjectival expressions:

$$[[\text{so tanzen ‘dance like this’}]] = \lambda e. \text{dance}(e) \& \text{SIM}(e, t, \mathcal{F})$$

[[*so groß ‘this tall’*]] =  $\lambda x. \text{SIM}(x, t, \mathcal{F}(\text{height}))$  where  $\mathcal{F}(\text{height})$  is meant to restrict the representation to the height dimension.

is quantitative (distance-based) (see Sect. 4.2).<sup>9,10</sup> Even more importantly, unlike Gärdenfors' conceptual spaces, multi-dimensional attribute spaces in the Umbach and Gust framework are integrated into referential semantics by means of generalized measure functions mapping referents to points in multi-dimensional attribute spaces. Note that this is just a generalization of degree semantics (e.g. Kennedy 1999) from the one-dimensional to the multi-dimensional case.

### 3 Three Types of Similarity Expressions

In this section the three types of similarity expressions—*so/such*, *ähnlich/similar* and *gleich/same*—will be compared focusing on semantic characteristics (for lexical and distributional data see Umbach 2014). First, *ähnlich/similar* as well as *gleich/same* are relational adjectives comparing two individuals. The second argument may be explicit (*Ann's car is similar to Berta's car*) or anaphoric (*Ann's car is similar*).<sup>11</sup> In contrast, *so/such* are demonstratives (to be used deictically as well as anaphorically). Even though the target of the demonstration (or antecedent) is not identical to the referent of the phrase—the referent of *such a table* is not (necessarily) identical to the table pointed to—it would be misleading to think of *so/such* as expressions relating two distinct individuals. This is obvious when considering reciprocal readings which are licensed by *ähnlich/similar* (as well as *gleich/same*), but not by *so/such* (*Anna and Berta have similar cars/\*have such cars*). Instead, these demonstratives create an ad-hoc set of items similar to the target—a set of tables similar to the table pointed to—which is then used to introduce a novel discourse referent (note that *so/such* are incompatible with definite determiners, *\*so der Tisch/\*such the table*).

Furthermore, while *ähnlich/similar* as well as *gleich/same* are predicates denoting pairs of individuals and may vary across indices, *so/such* are demonstratives. They refer directly to the target pointed at and block indexical shift (Kaplan 1989). This is shown in (8): (8a) is clearly true. But even though Adam and Ben both drive a Porsche,

<sup>9</sup>Voronoi tesselations are restricted to distance-based accounts with prototypes.

<sup>10</sup>Sassoon (2013) investigates the meaning of multidimensional adjectives such as *healthy* and *sick*. She suggests a classification by the way dimensions are combined presupposing that for each dimension there is some standard. *Conjunctive* adjectives require entities to reach the standard in all of their dimensions while *disjunctive* adjectives require the same for some dimensions. Comparatives are analyzed by means of counting dimensions. Sassoon's account is directed at the issue of dimension integration. Questions of similarity and indistinguishability do not play a role in her account.

<sup>11</sup>We ignore reciprocal and NP-dependent occurrences, as in *Anna has similar dogs./Anna and Berta have similar dogs*, see Beck (2000) on the meaning of *different*.

(8b) is false because the counterfactual index is irrelevant to the target of the demonstration—the speaker is still pointing to the old VW. In contrast, *ähnlich/similar* (as well as *gleich/same*) are evaluated at the counterfactual index, and thus (8c) is true.<sup>12</sup>

- (8) (Adam and Ben both drive a Porsche cabrio. Chris has an old VW. The speaker points to the car parked in front of the garden gate.)
- (scenario 1: Adam's car is parked in front of the gate)  
Ben hat auch so ein Auto/Ben has such a car, too. true
  - (scenario 2: Chris' car is parked in front of the gate)  
Wenn Adam vor dem Tor parken würde, hätte Ben auch so ein Auto./  
If Adam were parked in front of the gate, Ben would have such a car,  
too. false
  - (scenario 2: Chris' car is parked in front of the gate)  
Wenn Adam vor dem Tor parken würde, wäre Ben's Auto dem Auto vor dem Tor  
ähnlich. /  
If Adam were parked in front of the gate, Ben's car would be similar to the one in  
front of the gate. true

Another difference between the three types of similarity expressions is given by the selection of the dimensions of comparison. In the case of *so/such*, dimensions are first of all determined by the lexical meaning of the noun—dimensions to be considered for something to be a table or be a bike. Other dimensions can be relevant as long as they relate to properties suited to create a subkind of the kind corresponding to the noun. Take the noun *bike*. For something to be *such a bike* it has to be similar to the bike pointed at in relevant bike dimensions. There may be additional dimensions which are not specific for bikes, surfacing in properties like *rusty* or *dented*. But properties like *bought last year from her neighbor* or *fantastic* would not qualify for comparison. This is why the *namely* continuations in (9a) and (b) are unmarked whereas in (c) and (d) they are clearly bad. In the case of *so/such*, dimensions of comparison are not restricted to those determined by the lexical meaning of the noun, but they must not relate to indexical (in a broad sense) or evaluative properties, because indexical and evaluative properties are unsuited to create subkinds (experimental evidence is described in Umbach and Stolterfoht in prep., see also König and Umbach 2018, Sect. 5).

---

<sup>12</sup>Regarding ex. (8b), it could be objected that, in German, an equative construction would yield a true proposition—*Wenn Adam vor dem Tor parken würde, hätte Ben auch so ein Auto wie das vor dem Tor.* ('If Adam were parked in front of the gate, Ben would have a car like the one in front of the gate.') This effect is due to the fact that *so* in equatives is not a demonstrative but instead a correlative and does not refer at all.

- (9) a. Anna has a mountain bike. Berta has such a bike, too (namely a mountain bike).  
 b. Anna's bike is rusty and dented. Berta has such a bike, too (namely a rusty and dented one).  
 c. Anna has a bike bought last year from her neighbor. Berta has such a bike, too (#namely one bought last year from her neighbor).  
 d. Anna has a good bike. Berta has such a bike, too (#namely a good one).

Selection of dimensions is different in the case of *ähnlich/similar*. Consider the example in (10). First, while *so/such* phrases are perfect as kind-denoting terms in generic sentences, *ähnlich/similar* phrases are not, see (10a, b). Secondly, changing the (unacceptable) generic sentences in (10b) into the episodic sentence in (10c) reveals a clear difference in meaning: *so ein Geschenk/such a present* is something rare and valuable which can reasonably be considered as showing appreciation for the guest. A Panda bear serves this purpose, but an old manuscript or painting would do as well. In contrast, *ein ähnliches Geschenk/a similar present* need not be valuable, but it has to be similar to a Panda bear. When asked, what a similar present could be, informants mention tigers, rhinos, crocodiles etc. This is strong evidence that the *ähnlich/similar* version of similarity selects dimensions made salient by the antecedent.

- (10) (The prime minister received a Panda bear from the Chinese government.)  
 a. So ein Geschenk zeigt die Wertschätzung des Gasts./  
     Such a present demonstrates appreciation for the guest.  
 b. # Ein ähnliches Geschenk zeigt die Wertschätzung des Gasts. /  
     # A similar present demonstrates appreciation for the guest.  
 c. Ein ähnliches Geschenk brachte ihm im Vorjahr Kritik im eigenen Land ein. /  
     A similar present evoked protests in his own country last year.

In the case of *gleich/same*, there is a type and a token interpretation (Nunberg 1984). (11) may mean that Anna and Berta drive cars of the same type, or that Anna and Berta share a car (token).<sup>13</sup> The token interpretation yields referential identity,  $x = y$ , but the type interpretation is, first of all, just similarity—being indistinguishable with respect to dimensions given by the lexical meaning of the noun. Different from *so/such* and *ähnlich/similar*, additional dimensions are blocked for *gleich/same*. Suppose that Anna drives a Ford Fiesta. Then *the same car* on a type interpretation has to be a Ford Fiesta. But even if Anna's car is rusty and dented, *the same car* could

---

<sup>13</sup>There are prescriptive efforts to restrict German *gleich* to type readings and require token readings to be expressed by *selb*, but German speakers don't follow this rule. That does not imply, however, that there is no differences between *gleich* and *selb*, just that the rule is not descriptively correct, see Umbach (2019). Moreover, there are reasons to assume that the parallelism between German *gleich* and English *same* breaks down when it comes to type identity, in that *same* is closer to *selb* than to *gleich*.

be spotless. Obviously, non-car-specific dimensions like conditions of usage are irrelevant. Moreover, while *such a car* may deviate from the values of the antecedent in some dimensions—e.g. by having two instead of four doors—the same car has to be exactly like the antecedent in every car dimension.

- (11) Anna fährt das gleiche Auto wie Berta. /  
Anna drives the same car that Berta drives.

We will assume that for every noun there is a lexically associated canonical set of dimensions (called N-related dimensions). They are provided by criteria of application—what it means to be a table—and are not to be mistaken for criteria of identity.<sup>14</sup> Our hypothesis on the selection of dimensions of comparison is this<sup>15</sup>:

- (12) (i) *so/such* require a subset of the N-related dimensions to be considered and allow for additional dimensions as long as they are suited for kind-formation (see above). (Leeway in the values of measure functions depends on the given set of classifiers).
- (ii) *ähnlich/similar* require a set of dimensions made salient by the antecedent.
- (iii) *gleich/same* (type reading) require all and only N-related dimensions to be considered and measure functions to yield the same values:  $\mu(x)=\mu(y)$ . (Since the token reading denotes referential identity, dimensions are irrelevant.)

Let us finally consider reflexivity. In the example in (13) *so eine Feuerwehr/such a fire brigade* in (a) is anaphorically related to the previously mentioned team of fire fighters, which is the team the mayor intends to praise. So the referent of the *so/such phrase* is identical to the antecedent. When *so/such* is substituted by *ähnlich/similar*, as in (b), the mayor seems to praise a fire brigade different from the successful team, which appears strange in this context. A similar effect is found with *gleich/same*—(c) again gives the impression that there is another fire brigade (for (d) see below).

---

<sup>14</sup>Gupta (1980) postulates that nouns provide criteria of identity determining the way objects are counted (in addition to criteria of application). His famous example is

(a) Easyjet served 10 million passengers last year.  
(b) Easyjet served 10 million people last year.  
(a) can be true and (b) false at the same time because one person may count as two passengers on two different flights. Barker (2010) argues against this idea, attributing the effect to the fact that deverbal nominals like *passenger* may (but need not) give rise to a per-event reading in addition to the regular per-individual reading. The slightly absurd dialog below confirms Barker's position:  
On a flight to Bilbao in June 2017.

Flight attendant A: Look at seat 12a. This is the same passenger that flew to Barcelona in April 2016.

Flight attendant B: No, it is the same person but not the same passenger.

<sup>15</sup>Regarding ex. (12iii). Type identity of *gleich* may in addition be limited to mass produced entities and clones (Stephanie Solt p.c.).

- (13) (Mayor Friedmann expresses his gratitude towards the fire fighters, pointing out that it is a great achievement of the team that the fire did not flash over to the adjacent buildings. He says:)
- a. Wir in der Gemeinde freuen uns, dass wir so eine Feuerwehr haben! / We are happy to have such a fire brigade in our community!
  - b. Wir in der Gemeinde freuen uns, dass wir eine ähnliche Feuerwehr haben! / We are happy to have a similar fire brigade in our community!
  - c. Wir in der Gemeinde freuen uns, dass wir die gleiche Feuerwehr haben! / We are happy to have the same fire brigade in our community!
  - d. Wir freuen uns, dass wir die gleiche Feuerwehr wie die vor 10 Jahren haben! / We are happy to have the same fire brigade as the one 10 years ago!

(13a) clearly shows that in the case of *so/such* similarity is reflexive. (13b) shows that in the case of *ähnlich/similar* reflexive pairs are excluded. But we started out from the idea that the three varieties of similarity expressions are based on one common similarity relation—it would be unintuitive to have an irreflexive similarity relation *SIM'* in addition to the ‘regular’ reflexive one. More importantly, (13c) shows the same effect as in (13b): there seem to be two distinct fire brigades. It would be absurd, however, to claim that *gleich/same* are not reflexive. We will therefore postulate distinctiveness as a precondition of usage (due to the two-place character of the lexical items).<sup>16</sup>

Postulating distinctiveness as a precondition yields the required result for *ähnlich/similar*. Note, however, that in the case of *gleich/same* the distinctiveness effect is slightly different from what was found for *ähnlich/similar*. (13c) is strange only if there is no different description of the fire brigade available. But if the mayor earlier in his speech mentioned the fire brigade the community had 10 years ago, he could refer to the actual one by “the same fire brigade [as 10 years ago]” in the sense of token identity (suppose the group of fire fighters did not change), see (13d). So *gleich/same* do not require distinct referents but instead distinct senses—*Arten des Gegebenseins*—as in Frege’s distinction between sense and reference. The sentence *The morning star is the same star as the morning star* is decidedly odd whereas *The morning star is the same star as the evening star* is fine, which led Frege to distinguish sense and reference (Frege 1892). Accordingly, (13d) is fine because although the fire brigade referent is identical to the one 10 years ago (on the token reading) there are two different senses—*fire brigade now, fire brigade 10 years ago*.

Therefore, while *ähnlich/similar* presuppose distinctiveness of referents, *gleich/same*—on the token reading!—require distinctiveness of descriptions, or ways of identification. The type reading of *gleich/same*, on the other hand, requires that referents are distinct, which is trivial because otherwise it would not be a type reading.

---

<sup>16</sup>In Umbach (2014) *ähnlich* was said to carry a distinctiveness constraint, thereby explaining that additive particles appear redundant with *ähnlich* but not with *so* (... *Berta has such a car, too./?? a similar car, too*). But distinctiveness was wrongly conflated with irreflexivity in that paper.

Summing up, all of the three variants of similarity expressions can be analyzed as being based on a single similarity relation,  $\text{SIM}(x, y, \mathcal{F})$ . Their differences are due to differences in selecting dimensions of comparison and in different preconditions of usage.

- (14) a.  $[[\text{so/such}]] = \lambda P \lambda x. \text{SIM}(x, t, \mathcal{F}) \& P(x)$  where  $t$  is the target of demonstration  
and  $\mathcal{F}$  complies with (12) (i)
- b.  $[[\text{ähnlich/similar}]] = \lambda P \lambda y \lambda x. \text{SIM}(x, y, \mathcal{F}) \& P(x)$  where  $x \neq y$  and  $\mathcal{F}$  complies with (12) (ii)
- c.  $[[\text{gleich / same}]] =$   
type reading:  $\lambda P \lambda y \lambda x. \text{SIM}(x, y, \mathcal{F}) \& P(x)$  where  $x \neq y$  and  $\mathcal{F}$  complies with (12) (iii)  
token reading:  $\lambda P \lambda y \lambda x. x=y \& P(x)$  where  $x$  and  $y$  are given  
by different ways of identification

Two remarks: First, we do not touch upon the issue of constraints on determiners due to reasons of space, (see Umbach 2014). Secondly, the precondition of usage in (b) may be formulated as a presupposition. This is not possible in (c) because *way of identification* is an intensional notion, which is not (yet) available in the similarity framework (see Appendix).

## 4 Gradability of *ähnlich/similar*

This section focuses, first, on the question of how *ähnlich/similar* compares to other gradable predicates, and what it means for two items to be more similar than some other two items. In the second part of this section, cognitive models of similarity are considered from the point of view of gradability, and the basic ideas of the model suggested in this paper are introduced (technical details are given in the Appendix). Finally, we will give a tentative answer to the question of why *ähnlich/similar* are gradable but neither *so/such* nor *gleich/same* are.

### 4.1 What Does It Mean to Be More Similar?

For relative gradable adjectives, the truth of the positive form depends on the relevant comparison class—*Anna is tall* may be true when comparing Anna to her classmates and false when comparing her to her basketball teammates. Absolute gradable adjectives do not require comparison classes because they make use of minimal or maximal degrees of the gradable property—*The door is closed* is true only if it is maximally closed, and false if it is ajar (cf. Kennedy and McNally 2005). So unlike relative adjectives, absolute ones include a lower or upper bound (or both).

Neither *ähnlich* nor *similar* admit reference to overt comparison classes, see (15a). The examples improve slightly when referring to a relativizing state of affairs, see

(15b). Examples are unmarked when referring to dimensions of comparison (15c), which is no surprise since similarity generally requires dimensions.

- (15) a. ??? Für ein ärmelloses Sommerkleid ist Annas Kleid dem von Berta ähnlich. /  
??? For a sleeveless summer dress Anna's dress is similar to Berta's dress.
- b. (?) Dafür dass es aus den Sechzigern stammt, ist Annas Kleid dem von Berta ähnlich. /  
(?) Taking into account that it is from the sixties Anna's dress is similar to Berta's dress.
- c. Im Hinblick auf Schnitt und Material ist Annas Kleid dem von Berta ähnlich. /  
Anna's dress is similar to Berta's dress with respect to cut and fabric.

Maxima can be linguistically indicated with the help of degree modifiers like *vollständig* and *completely*. As noted earlier, neither *ähnlich* nor *similar* admit these modifiers.<sup>17</sup> In fact, the combinations *vollständig ähnlich* and *completely similar* appear inconsistent, see (16a). Intuitively, if two items are similar, they do not fully agree in their properties, and if agreement is complete, the items are no longer called *ähnlich/similar* but instead *gleich/same*. So there is an upper bound, a maximum at which two items cannot possibly be more similar than they are. But this maximum is denoted by *gleich/same*, on either a token or a type reading, see (16b).<sup>18</sup>

- (16) a. ?? Anna fährt ein vollkommen ähnliches Auto wie Berta./  
?? Anna drives a car completely similar to Berta's.
- b. Anna fährt das gleiche Auto wie Berta. /  
Anna drives the same car that Berta drives.

The intuition that *gleich/same* denote maximal similarity is based on the idea that the more features two items share, the more similar they are.<sup>19</sup> It is important to note, however, that this is one of two opposite perspectives. If there is a fixed set of features, then two items are more similar than two other items if they share more of these features.<sup>20</sup> If, on the other hand, the set of features is variable, then two items

<sup>17</sup>Corpus research for *completely similar* in COCA (more than 500 million words) returned three tokens; *vollständig ähnlich* in DEWAC (more than 1 billion words) returned only one token. A few more were found for *vollkommen ähnlich* und *völlig ähnlich*, the latter including a famous subtitle of a drawing showing Leibniz in Park Herrenhausen saying

*Leibniz behauptet, daß nicht zwei Blätter einander völlig ähnlich seien.*

*Leibniz claims that no two leaves are completely similar.*

<http://www.akg-images.de/archive/Leibniz-behauptet--da%C3%9F-nicht-zwei-Blatter-einander-vollig-ahnlich-seien-2UMDHUKPV6X.html>

<sup>18</sup>As one reviewer noted, this behavior is analogous to open intervals since the margin is not contained but we can come arbitrarily close.

<sup>19</sup>We use an informal notion of ‘feature’ here, like ‘property’, or ‘dimension + value’.

<sup>20</sup>This is the perspective in Tversky (1977).

may be similar w.r.t. a reduced feature set, even if they were not similar in the original set. Take lens resolution in a camera, which is responsible for the details that can be distinguished. If lens resolution is given, similarity can only be increased by changing the facts in the world. But if lens resolution is decreased similarity is increased in the sense that two items may be similar even if they were not similar in the original resolution (while facts in the world did not change). The second perspective is the one taken in the next section.

Considering *gleich/same* from this perspective, both the token and the type reading entail maximal discriminating capacity in the following sense: The type reading implies similarity, i.e. indistinguishability, in any representation spanned by N-related dimensions regardless how fine-grained it might be, and the token reading implies similarity in any representation at all (i.e. including accidental properties).

## 4.2 *Gradability and Granularity*

In Cognitive Science, models of similarity are either distance-based or feature-based. Distance-based models, for example Gärdenfors' (2000) *Conceptual Spaces*, start out from distances between points in a geometrical space representing objects of the domain in question. Similarity is determined by distance—the closer the points are (in a given metric) the more similar are the corresponding objects. Similarity is an intrinsic component of geometric representations and is exploited, e.g., in defining convexity.

In a distance-based model the notion of distance provides a “degree” of similarity. In degree-based accounts of gradability the meaning of the comparative, say, *taller*, is given by comparing degrees—*a is taller than b* iff a's height exceeds b's height. The positive, *tall*, is defined on top of the comparative by making use of a threshold provided by a comparison class (e.g. Bierwisch 1987; Kennedy 1999)—*a is tall* iff a's height exceeds the threshold of the relevant comparison class.<sup>21</sup>

The comparative of *ähnlich/similar* can be straightforwardly defined in distance-based models via the notion of distance (see, e.g., the comparative semantics for resemble in Meier 2009). The problem would be the positive. It is hard to imagine a way to define a predicate *similar* on the basis of the comparative, because there is no principled way to determine the threshold—what would be a plausible distance for two tables to count as similar?

---

<sup>21</sup>For a degree-based account of *similar/different* see also Alrenga (2007).

The other type of Cognitive Science models of similarity are feature-based ones, most prominently Tversky's (1977) *contrast model*. Tversky argued that there are empirical findings in conflict with the basic axioms of metric distance functions<sup>22</sup>: (a) minimality is problematic in view of results concerning the identification probability for identical stimuli, (b) symmetry is apparently false—the judged similarity of North Korea to Red China exceeds the judged similarity of Red China to North Korea—and (c) 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.<sup>23</sup>

In view of these issues Tversky claimed that "... the assessment of similarity between stimuli may better be described as comparison of features rather than as the computation of metric distance between points" (p. 328). He proposed a model in which similarity between two objects is computed on the basis of common and distinctive features: Similarity of two objects increases with an increase of common features and/or a decrease of distinctive ones.<sup>24</sup> This idea is modelled by a function  $S$  taking weighted sums of the feature sets  $A$  and  $B$  of objects  $a$  and  $b$  to an interval scale such that  $\text{sim}(a, b) \leq \text{sim}(c, d)$  iff  $S(a, b) \leq S(c, d)$ , where  $S(a, b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$ .<sup>25</sup>

As before in distance-based models, the notion of similarity in Tversky's feature-based model corresponds to a "degree" of similarity, thereby facilitating comparative statements. And as before, it is hard to imagine a way to define a predicate *similar* on the basis of the comparative because there is no principled way to determine the threshold.

The account of similarity proposed in this paper is feature-based. But instead of summing up common and distinctive features it makes use of dimensions and of classifiers determining whether values on these dimensions count as distinct. Similarity is defined in this account as indistinguishability with respect to given dimensions and classifiers: Two objects are similar if relative to the relevant dimensions and classifiers they are indistinguishable (see Appendix). In this account, the positive form *similar* is given, and the comparative form, *more-similar*, has to be defined on the basis of the positive.

<sup>22</sup>A metric distance function  $\delta$  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 ( $\delta(a, b) + \delta(b, c) \geq \delta(a, c)$ ).

<sup>23</sup>It has to be mentioned though that these results are highly controversial. Before dismissing transitivity on the basis of the Jamaica/Cuba/Russia example, one should consider the role of switching features within the two comparison steps. On symmetry, there is a detailed study by Gleitman et al. (1996) showing that the alleged asymmetry hinges on the way of presentation. In Tversky's original studies presentation was directional (*North Korea is similar to Red China*). As soon as presentation is non-directional (*North Korea and Red China are similar*) similarity is found to be symmetric (which was already suggested by Tversky himself). For reflexivity, see the discussion in Sect. 3.

<sup>24</sup>When speaking of features, Tversky refers to what we would call dimension + value pairs, that is properties.

<sup>25</sup> $\alpha, \beta, \theta$  denote weighting functions and  $f$  denotes a nonnegative scale.

In addition to degree-based accounts, there are so-called vague-predicate accounts of gradability, most prominently Klein (1980). In the latter, the comparative is defined on the basis of the positive form by making use of different interpretation contexts, i.e. (tripartite) partitions of the domain determining the extension of predicates. For example, *a is taller than b* is true if there is an interpretation context such that *a* counts as tall while *b* does not. The pros and cons of the two approaches have been the topic of a longstanding debate. One core issue is that degree semantics presupposes degrees, which are natural with adjectives like *tall* and *old*, since these adjectives are associated with units of measurement. But what would be degrees in the case of multidimensional adjectives like *skillful* and *good* and *ähnlich/similar*? If you think of multidimensional adjectives as spanning a multidimensional space, points in this space may be considered as degrees. But since points in a multidimensional space lack a natural order, some extra order has to be imposed (as, e.g., in Sassoon 2013, see footnote 10). This seems to suggest that in the case of multidimensional adjectives, vague-predicate approaches are more natural.

We adapt the idea of vague-predicate approaches by making use of *representations* of different granularity. Less granular representations have less discriminating capacity (pace dimensions and classifiers), and the lower the discriminating capacity of a representation is, the more items are similar, i.e. indistinguishable. Since the basic predicate *similar* is defined relative to a representation, the comparative will also be relative to a representation. We define the comparative in the following way:

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 a less granular representation  $\mathcal{F}'$  such that *a* and *b* are similar in  $\mathcal{F}'$  while *c* and *d* are not (Appendix, Definition 6, see also the remark on lens resolution at the end of Sect. 4.1).

Comparing this account to the Kleinian vague-predicate account, there are two points to be noted: First, one major characteristic of the Kleinian account is the elimination of degrees. However, the representations employed in defining a comparative of the similarity predicate include points in attribute spaces, which are in some sense analogous to degrees, thereby raising the question of why, in the similarity-based account, degree-like entities still play a role.<sup>26</sup> The answer is straightforward: Klein assumes predicates denoted by the positive forms, e.g. *tall*, to be given. The similar relation, in contrast, is not assumed to be given, but instead defined via representations. So points in attribute spaces are already required when defining the predicate denoted by the positive forms *ähnlich/similar*, independent of the definition of the comparative.

On a related issue, while Klein's account presupposes a natural order on the items in the domain, e.g., w.r.t height, there is no natural order of similarity—being similar is in general relative to a representation. The requirement for Kleinian interpretation contexts to be consistent with the order on the domain can be seen as a grounding requirement: Interpretations must comply with the given structure of the world. Representations are the counterpart to interpretation contexts, raising the question of

---

<sup>26</sup>Many thanks to the reviewer who pointed out this question.

whether there is a grounding requirement for representations. In fact, there is such a requirement built into the similarity framework by means of a consistency constraint: Classifiers have to be consistent with the results of the predicates they correspond to (Appendix, Definition 2).

So from a broader perspective, both representations and Kleinian interpretation contexts are grounded in factual matters. The Kleinian account directly refers to orderings in the domain—this is why interpretation contexts need not themselves be ordered. In the similarity account, representations have to be ordered, thereby lifting the Kleinian order requirement to the level of representations.

Let us finally come to the question why the *ähnlich/similar* variety of similarity expressions is gradable while neither *so/such* nor *gleich/same* are. It turns out that the explanation is straightforward, in both cases referring to the need of a less granular representation in defining the comparative.

In the case of *so/such*, representations other than the actual one are inaccessible because *so/such* are demonstratives instead of content words and thus have to be evaluated in the actual context. Since representations are clearly part of the context, they are part of what cannot be shifted in the case of demonstratives.

In the case of *gleich/same*, maximal discriminative capacity is required—type identity entails indistinguishability in any representation spanned by the N-related dimensions, token identity entails indistinguishability in any representation whatsoever. In either case, defining a comparative making use of less granular representations is ruled out.

## 5 Conclusion

In this paper, three types of expressions were compared that express similarity in some sense—*so/such*, *ähnlich/similar* and *gleich/same*—starting from the observation that *ähnlich/similar* are gradable but neither *so/such* nor *gleich/same* are. Their semantics was compared on the basis of a common similarity relation revealing differences in, e.g., the selection of dimensions of comparison and the status of reflexive pairs. The similarity relation is spelled out as indistinguishability in a mathematically precise framework of representations combining multi-dimensional attribute spaces with classification functions. A predicate *more-similar* was defined in a Kleinian style making use of representations of varying granularity. The definition predicts gradability of *ähnlich/similar* but not of *so/such* and *gleich/same*.

The paper provides a semantic analysis of three closely related types of expressions which have, if at all, been considered only in isolation. Moreover, it can be seen as a contribution to a long-standing debate on sameness and indistinguishability in natural language (see, e.g., Nunberg 1984, 2004; Lasersohn 2000; Barker 2010).

Future research will extend the analysis to include demonstratives like *dieser/this*, the notorious contrast between German *derselbe* and *der gleiche* and the contrast between English *same* and *identical*, and also include expressions of difference.

**Acknowledgements** We would like to thank two reviewers for their detailed and very helpful reviews. We would also like to thank the audience of the Cognitive Structures conference (Düsseldorf, 2016) and of the ZAS workshop “Records, Frames and Attribute Spaces” (Berlin, 2018), and in particular Robin Cooper, Louise McNally, Wiebke Petersen and Stephanie Solt for their valuable comments. Finally, we like to express our gratitude to the editors of this volume for their patience and support. The first author acknowledges financial support by the German Research Foundation (Deutsche Forschungsgemeinschaft), UM 100/1-3.

## Appendix: Granularity in Multi-dimensional Attribute Spaces

In the Appendix, the basic mathematical ideas and definitions of the similarity framework are presented. For more details see Gust and Umbach (to appear).

### *Domains and Representations*

The core of the appendix are sets of *representations* equipped with a preorder structure. This preorder implements a concept of granularity and will be used to construct a predicate *more\_similar* based on a similarity relation. We start with defining a *domain* as a subset of the universe 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:

- $D$  a set of individuals/entities (called the carrier of the domain),
- $P = \{p_1, \dots, p_n\}$  a set of predicates over  $D$ , (a subset of the powerset of  $D$ )
- $\_^+: P \rightarrow \wp(D)$  a function which assigns (a finite set of) positive examples to each predicate,  
for  $\_^+(p)$  we write  $p^+$
- $\_^-: P \rightarrow \wp(D)$  a function which assigns (a finite set of) negative examples to each predicate,  
for  $\_^-(p)$  we write  $p^-$
- $\forall p \in P : p^+ \cap p^- = \emptyset$

We view the elements of  $D$  as entities to which we have only indirect access via a (generalized) measure function  $\mu$  which constructs representations of the entities in  $D$  in an attribute space  $F$  much like observables in physics. Attribute spaces are common structures for representation.<sup>27</sup> They generalize vector space approaches in allowing heterogeneous dimensions equipped with value sets of different scales (nominal, ordinal, interval, ratio etc.), where value sets may themselves be attribute spaces.

---

<sup>27</sup>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).

An attribute space  $F$  is given by a set of attributes  $A = \{a_1 \dots a_n\}$ , 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  $V_{a1} \times \dots \times V_{an}$ , the carrier of the attribute space  $F$ .

A *representation* includes an *attribute space*  $F$ , a (generalized) *measure function*  $\mu$  mapping elements of a domain into the attribute space, and a set of *classification functions*  $p^*$  talking about points in the attribute space. These classification functions (short *classifiers*) serve as *approximations*<sup>28</sup> of the predicates in  $P$ . Moreover, the extensions of the classifiers will be assumed to be convex. This means that  $F$  comes with a convex closure operator  $cl$  and  $p^*$  must be *true* on  $cl(\mu(p^+))$ .<sup>29</sup> Intuitively, using the n-dimensional Euclidean space as an example, this means that 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  $F$ , with a closure operator  $cl$   
(we will write  $F$  for  $\langle F, cl \rangle$  if we are not interested in the closure operator  $cl$ )
- a measure function  $\mu: D \rightarrow F$ ,
- a function  $\_^*: P \rightarrow \{\text{true}, \text{false}\}^F$  (again we write  $p^*$  for  $\_^*(p)$  and call them classifiers)<sup>30</sup>

together with the consistency conditions

- $\forall p \in P$  the extension of  $p^*$  must be convex in  $\langle F, cl \rangle$
- $\forall p \in P \forall x \in p^+: p^*(\mu(x)) = \text{true}$
- $\forall p \in P \forall x \in p^-: p^*(\mu(x)) = \text{false}$

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

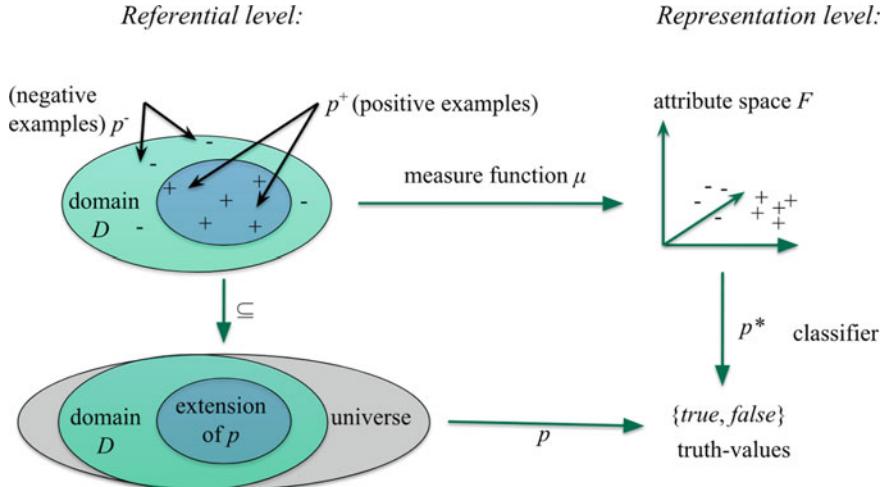
As mentioned above, attribute spaces are familiar methods of representation. What distinguishes attribute spaces and the representations proposed in this paper is the idea of classifiers on attribute spaces. On the worldly side, a domain includes a set of relevant predicates  $p \in P$ . On the representation 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, they have to agree in truth value on the set of positive/negative exemplars known for the original predicate (see Definition 2).

While attribute spaces can provide highly structured representations, classifiers provide binary features (attributes with possible values in  $\{\text{true}, \text{false}\}$ ). Given a set of basic classifiers we assume the possibility to construct derived classifiers by

<sup>28</sup>More precisely:  $p^* \circ \mu$  approximates  $p$ .

<sup>29</sup>This includes all points in the convex closure of the images of the positive exemplars. For convex closure operators see Korte et al. (1991). 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  $X$ .

<sup>30</sup>where  $\{\text{true}, \text{false}\}^F$  is the set of characteristic functions in  $F$ . Additionally we expect that classification functions come with algorithmic methods to compute these functions.



**Fig. 1** Domains and representations

logical operations: For the logical conjunction this works fine (convex sets are closed under intersection). For the logical disjunction we have to apply the convex closure operator  $cl$  to the result. For negation this does not work at all. So we do not allow to define complex classifiers by applying negation to elementary ones. We name the set of derived classifiers  $\tilde{P}^*$  (Fig. 1).

### Indiscernibility

Given a system of predicates  $P$  we can ask, which elements in a domain  $D$  can be distinguished. There are two reasons why we may not be able to distinguish between two elements of  $D$ :

- Two elements may lead to the same value of the function  $\mu$ . Then there is no way to discriminate between the two elements.
- The two elements disagree on  $\mu$  (so we see them as different), but they agree on all classification functions in  $\tilde{P}^*$ .

For the second case we borrow the term *indiscernible* from Rough Set Theory (Pawlak 1998):

#### Definition 3 Indiscernible

Given a representation  $\mathcal{F} = \langle F, \mu, \sim^*, \mathcal{D} \rangle$  with  $\mathcal{D} = \langle D, \sim^+, \sim^-, P \rangle$  we define:

For  $x, y \in F : x \sim_{\mathcal{F}} y \equiv \forall q \in \tilde{P}^*, q(x) \leftrightarrow q(y)$   
where  $\tilde{P}^*$  is the set of all derived classifiers.

This relation talks about points in  $F$ . However, the similarity relation we are interested in talks about elements of the domain  $D$ . So we have to apply the measure function first:

**Definition 4 Similar**

$$\forall x, y \in D : \text{sim}(x, y, \mathcal{F}) \equiv \mu(x) \sim_{\mathcal{F}} \mu(y)$$

Obviously, Definition 4 defines an equivalence relation on  $D$ .

The indiscernibility relation provides attribute spaces with a level of granularity, facilitating comparison of attribute spaces of distinct granularity which are identical otherwise.

### Granularity and Gradability

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. Since indiscernability of entities in a representation depends on the set of dimensions of the corresponding  $F$  and of the corresponding predicates  $P$  given in  $\mathcal{F}$ , granularity of representations depends on these parameters, too. Maybe there are more constraints we would like to impose on systems of representations to make such a system coherent in some sense. But we will not go into details here.

On representations we can define a reflexive and transitive relation (a preorder), which relates granularity levels:

**Definition 5 Coarser representation**

Given two representations

$$\begin{array}{ll} \mathcal{F} = \langle F, \mu, *_*, \mathcal{D} \rangle & \text{with } \mathcal{D} = \langle D, *_+, *_-, P \rangle \\ \mathcal{F}' = \langle F', \mu', *_*, \mathcal{D}' \rangle & \text{with } \mathcal{D}' = \langle D', *_+', *_-' , P' \rangle \end{array}$$

we define:

$\mathcal{F}'$  is coarser than  $\mathcal{F}$ ,  $\mathcal{F}' \geq \mathcal{F}$  iff

- (a) there is a function  $f$  such that  
the following diagram commutes:

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

$$\begin{array}{ccc} D' & \xrightarrow{\mu'} & F' \\ \uparrow \text{IU} & & \uparrow f \\ D & \xrightarrow{\mu} & F \end{array}$$

This means that what is indiscernible in the finer representation cannot be discriminated in the coarser representation. The strict version  $<$  is used such that  $\mathcal{F}$  is finer than  $\mathcal{F}'$ , or  $\mathcal{F}'$  is coarser than  $\mathcal{F}$ , if  $\mathcal{F} < \mathcal{F}'$  (in a preorder:  $x < y$  if  $x \leq y$  but not  $y \leq x$ ).

Based on our similarity relation  $\text{sim}$  and the preorder on representations we define a general relation  $\text{more\_sim}(a, b, c, d, \mathcal{F})$ , which 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 6 More similar**

Given a representation  $\mathcal{F}$  we define

$\text{more\_sim}(a, b, c, d, \mathcal{F})$  iff

- (a)  $\exists \mathcal{F}' \geq \mathcal{F} : \text{sim}(a, b, \mathcal{F}') \wedge \neg\text{sim}(c, d, \mathcal{F}')$
- (b)  $\forall \mathcal{F}' \geq \mathcal{F} : \text{sim}(c, d, \mathcal{F}') \rightarrow \text{sim}(a, b, \mathcal{F}')$

The widely used 3-place version  $\text{more\_sim}(x, y, z, \mathcal{F})$  in the sense that  $x$  is more similar to  $z$  than  $y$  can be defined straightforwardly by:

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

This approach shows how to model different similarity situations by selecting suitable sets of representations.

## References

- Alrenga, P. (2007). *Dimensions in the semantics of comparatives*. Dissertation, University of California, Santa Cruz.
- Anderson, C., & Morzycki, M. (2015). Degrees as kinds. *Natural Language & Linguistic Theory*, 33–3, 791–821.
- Barker, C. (2010). Nominals don't provide criteria of identity. In M. Rathert & A. Alexiadou (Eds.), *The semantics of nominalizations across languages and frameworks* (pp. 9–24). de Gruyter Berlin.
- Beck, S. (2000). The semantics of *different*. Comparison operator and relational adjective. *Linguistics & Philosophy*, 23(2), 101–139.
- Bierwisch, M. (1987). Semantik der Graduierung. In M. Bierwisch & E. Lang (Eds.), *Grammatische und konzeptuelle Aspekte von Dimensionsadjektiven* (pp. 91–286). Berlin: Akademie Verlag.
- Carlson, G. N. (1980). Reference to kinds in English. Garland, New York & London.
- Carpenter, B. (1992). *The logic of typed feature structures*. Cambridge Tracts in Theoretical Computer Science, Cambridge University Press.
- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift f. Philosophie u. phil. Kritik*, NF 100, S.25–50.
- Gärdenfors, P. (2000). *Conceptual spaces*. MIT Press.
- Gleitman, L., Gleitman, H., Miller, C., & Ostrin, R. (1996). Similar, and similar concepts. *Cognition*, 58, 321–376.
- Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and Projects* (pp. 437–447). Indianapolis and New York: The Bobbs Merrill Company.
- Gupta, A. (1980). *The logic of common nouns: an investigation of quantified modal logic*. New Haven, CT: Yale University Press.
- 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.
- Gust, H., & Umbach, C. (to appear) A qualitative similarity framework for the interpretation of natural language similarity expressions. In L. Bechberger, K. -U. Kühnberger, & M. Liu (Eds.), *Concepts in action - representation, learning and application. Language, cognition, and mind*. Springer.

- 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*. New York: Garland Press.
- Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, 81, 345–381.
- Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and 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.
- Lasersohn, P. (2000). Same, models and representation. In *Proceedings of semantics and linguistic theory* (Vol. 10).
- Meier, C. (2009). A comparative analysis for resemblance. In L. Kálmán (Ed.), *Proceedings of the 10th Symposium on Logic and Language, Hungarian Academy of Sciences*, pp. 141–148.
- Minsky, M. L. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision*. New York.
- Nunberg, G. (1984). Individuation in context. In *Proceedings of the West Coast conference on formal linguistics* (pp. 203–217). Stanford University.
- 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).
- Sassoon, G. (2013). A typology of multidimensional adjectives. *Journal of Semantics*, 30, 335–380.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352.
- Umbach, C., & Gust, H. (2014). Similarity demonstratives. *Lingua*, 149, 74–93.
- Umbach, C. (2014). Expressing similarity: On some differences between adjectives and demonstratives. In *Proceedings of IATL 2013, MIT working papers in linguistics*.
- Umbach, C. (2019) *Dasselbe oder das gleiche? - warum wir dann doch dieselbe Pizza bestellen können wie der Gast nebenan*. Jahrbuch 2018 des Zentrums Allgemeine Sprachwissenschaft (ZAS), Berlin.
- Umbach, C., & Stolterfoht, B. (in prep.). *Ad-hoc kind formation by similarity*.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# **Cognition and Psychology**

# Escitalopram Restores Reversal Learning Impairments in Rats with Lesions of Orbital Frontal Cortex



David S. Tait, Ellen E. Bowman, Silke Miller, Mary Dovlatyan,  
Connie Sanchez, and Verity J. Brown

**Abstract** The term ‘cognitive structures’ is used to describe the fact that mental models underlie thinking, reasoning and representing. Cognitive structures generally improve the efficiency of information processing by providing a situational framework within which there are parameters governing the nature and timing of information and appropriate responses can be anticipated. Unanticipated events that violate the parameters of the cognitive structure require the cognitive model to be updated, but this comes at an efficiency cost. In reversal learning a response that had been reinforced is no longer reinforced, while an alternative is now reinforced, having previously not been (A+/B— becomes A-/B+). Unanticipated changes of contingencies require that cognitive structures are updated. In this study, we examined the effect of lesions of the orbital frontal cortex (OFC) and the effects of the selective serotonin reuptake inhibitor (SSRI), escitalopram, on discrimination and reversal learning. Escitalopram was without effect in intact rats. Rats with OFC lesions had selective impairment of reversal learning, which was ameliorated by escitalopram. We conclude that reversal learning in OFC-lesioned rats is an easily administered and sensitive test that can detect effects of serotonergic modulation on cognitive structures that are involved in behavioural flexibility.

---

D. S. Tait · E. E. Bowman · V. J. Brown (✉)

School of Psychology and Neuroscience, University of St. Andrews, St. Mary’s College, South Street, St. Andrews, Fife KY16 9JP, Scotland, UK

e-mail: [vjb@st-andrews.ac.uk](mailto:vjb@st-andrews.ac.uk)

D. S. Tait

e-mail: [dst@st-andrews.ac.uk](mailto:dst@st-andrews.ac.uk)

S. Miller · M. Dovlatyan · C. Sanchez

Lundbeck Research USA, Inc., 215 College Rd, Paramus, NJ 07652, USA

e-mail: [Silke.Miller@sagerx.com](mailto:Silke.Miller@sagerx.com)

M. Dovlatyan

e-mail: [mary.dovlatyan@gmail.com](mailto:mary.dovlatyan@gmail.com)

C. Sanchez

e-mail: [Connie.SanchezMorillo@alkermes.com](mailto:Connie.SanchezMorillo@alkermes.com)

**Keywords** Goals · Free-will · Cognitive structures · Introspection · Cognitive flexibility · Rats · Reversal learning

## 1 Introduction

The frontal lobes of the human brain are thought to be the ‘seat of being’, providing functions that are quintessentially human. These include language but also functions related to having goals, considering consequences, weighing options, abstracting rules, making plans for the future, and free-will: in short, the frontal lobes hold the cognitive structures that give rise to the essence of human ‘self’. These are what Whitehead invoked when he wrote “*the life of a human being receives its worth, its importance, from the way in which unrealised ideals shape its purposes and tinge its actions. The distinction between men and animals is in one sense only a difference in degree. But the extent of the degree makes all the difference*” (Whitehead 1938, pp. 37–38).

It is not far-fetched to suggest that a hungry foraging rat has ‘unrealised ideals’ and that these are brought to bear in driving its behaviour and response choice, which determine future action. Furthermore, the frontal lobes of the rat contribute to this goal-directed behaviour and, from this, cognitive structures may be inferred. Therefore, quantifying this behaviour should demonstrate that it is possible, even if only within a relatively restricted cognitive domain, to measure the extent of the degree of difference referred to by Whitehead (1938).

Humans can verbalise many mental (cognitive) functions by introspection and communicate this to others. Without recourse to language, however, cognition cannot be directly measured, but rather only indirectly inferred from behaviour. The challenge then becomes that of finding suitable measures of behaviour that reflect the cognitions of interest in different species in order to take a comparative approach to understanding the neural basis of cognition. Such an approach has the obvious value that it could inform our understanding of fundamental properties of cognitive operations (Miller and Cohen 2001). However, there is an additional potential benefit, in that it enables the refinement of ‘animal models’ of human psychiatric disorders, such as schizophrenia or depression, in which cognitive flexibility is impaired (Murray et al. 2008; Kehagia et al. 2010; Murphy et al. 2012; Gilmour et al. 2013; Waltz 2017). In recent years pharmaceutical companies have curtailed investment in, or abandoned altogether, research into treatments for mental illness and other funders are not stepping in to counteract this trend. We recently argued that one of the reasons for this retreat is that ‘translational research’ has often failed to deliver its promise but, while limits of ‘animal models’ must be acknowledged, they do have value in providing an understanding of the neural mechanisms of specific symptoms (Insel et al. 2012).

Thus, there are multiple good reasons to identify those cognitive structures that are relevant for human health and wellbeing and are both likely to be evolutionarily conserved and can be readily measured and quantified in different species.

The capacity to behave flexibly is an adaptation that is fundamental for evolutionary fitness and is quantifiable in many different species. This makes studies of behavioural, and the presumed underlying cognitive, flexibility exemplary for this purpose.

## ***1.1 How Is Behavioural Flexibility Measured and Cognitive Flexibility Inferred?***

Cognitive structures improve the efficiency of information processing by providing a situational framework within which there are parameters governing the nature and timing of information and appropriate responses can be anticipated. In a highly predictable situation, unanticipated events require flexibility: the cognitive model is updated so that appropriate responses are generated. However, this updating incurs a cost, usually measured as additional time or experience required to learn under the changed conditions.

Most assays of cognitive flexibility exploit paradigms from the early psychology literature measuring perceptual attentional shifting (examples include the Wisconsin Card Sorting Test (Berg 1948) and the intra-/extra-dimensional (ID/ED) set shifting task (Lawrence 1949) or response switching (examples include task switching (Jersild 1927) and ‘learning set’ (Harlow 1949)). Some tests include elements of both perceptual shifting and task or response switching (see Floresco and Jentsch 2011), which could be problematic if shifting and switching are separable processes (for an excellent discussion of this see Ravizza and Carter 2008)). The third paradigm that is frequently used as a presumed measure of cognitive flexibility is reversal learning: after one reward pairing has been learned (e.g., ‘A+/B−’) it is reversed (e.g., ‘A−/B+’). Reversal learning has a long history of use, but it has become increasingly popular, particularly in the last decade, because of the ease with which it can be measured in different species, making it particularly useful for translational research (for review, see Izquierdo et al. 2017).

In all of these measures of cognitive flexibility, the assumption is that a cognitive structure is formed due to the repetition of a particular situational context (i.e., a stable ‘A+/B−’ association; an attentional focus on a particular stimulus feature; an effective response strategy). The anticipation of future stability means that when it is violated (i.e., ‘A+/B−’ becomes ‘A−/B+’; another stimulus attribute is relevant; an alternative response strategy is more effective), there is a ‘cost’, measured in retardation of learning, as the cognitive model is updated.

It has long been established that reversal learning is more rapid if the reversal is a reversion to a previous learned association. Furthermore, reversals are particularly rapid when they repeat serially (Harlow 1949). The benefit from repeating a reversal could arise in part from familiarity with the particular stimuli and the task requirements and is thus similar to the benefit of over-training (Dhawan et al. 2019). A benefit of repeating a reversal could also be due to incorporation into the cognitive structure

the concept that ‘reversals may occur’ (Izquierdo et al. 2017). In this study, we sought to tease these apart in the context of lesions of the orbital frontal cortex (OFC). We selected this particular brain region because it has repeatedly been shown to impair reversal learning in many different forms (for review see Izquierdo et al. 2017). In addition, serotonin has been implicated in reversal learning (Boulougouris et al. 2007; Bari et al. 2010; Brigman et al. 2010). Therefore, we investigated the effects of the selective serotonin reuptake inhibitor (SSRI), escitalopram, on discrimination and reversal learning in OFC-lesioned rats, and on prefrontal Fos immunoreactivity.

## 2 Methods

### 2.1 Animals

Twenty-eight naïve male Lister hooded rats (Harlan, UK) were used. The rats were pair-housed and maintained on a 12-h light/dark schedule (lights on at 7 a.m.), with a diet of 15–20 g of standard laboratory chow each day with water available ad libitum. The initial weight range was between 300 and 350 g. At completion of the experiment the weight range was between 310 and 390 g. All procedures were carried out in accordance with the UK Animals (Scientific Procedures) Act 1986.

### 2.2 Apparatus

The apparatus for the task and the basic testing protocol was the same as used during the rat attentional set-shifting task and have been described in detail elsewhere (Birrell and Brown 2000; Tait et al. 2018). In brief, the testing arena was constructed from large plastic home-cages (69.5 cm × 40.5 cm × 18.5 cm), with internal wooden runners permitting Perspex panels to selectively occlude either or both of two adjacent compartments, occupying one-third of the length of the cage, from the waiting area (the remaining two-thirds of the length). Within each of these compartments a ceramic digging bowl, containing scented digging media, could be placed.

### 2.3 Surgery

Fourteen rats were anaesthetised with an isoflurane (4% and reduced to 1% to maintain anaesthesia) and oxygen mix. 0.06 M ibotenic acid was administered bilaterally using a 0.5 µl Hamilton syringe with a 30 gauge needle attached, to target the orbital frontal cortex, at stereotaxic co-ordinates; tooth bar –3.3 mm, AP +4.0 mm, ML ±2.0 mm, DV –4.5 mm (from skull surface) (0.3 µl per site) over 2.5 min. The needle

was left in situ for 3 min after administration. Rats were administered a 0.05 ml injection (s.c.) of the anti-inflammatory, carprofen, and a 0.25 ml injection (i.p.) of the sedative, diazepam, prior to surgery. One lesioned rat died two weeks post-surgery, and before any testing.

Fourteen rats were administered sterile phosphate buffer instead of ibotenic acid and were assigned to the control groups.

## **2.4 Experiment 1: The Effects of Escitalopram on Reversal Learning**

### **2.4.1 Behavioural Training**

Between 10 and 20 days after surgery, 11 rats (lesion group n = 5; control group n = 6) were tested on the reversal learning task. The rats were first given experience of digging in ceramic bowls (of the size used for the test) and habituating to the food reward. Bowls were placed in the home-cage, filled with sawdust and a quantity of Honey Loops® (Kellogg Company, Manchester, UK). By the following morning, the food was always eaten. On the training day, rats were placed in the waiting areas of the testing cage, and underwent three stages of training. In stage 1, sawdust-filled bowls, with food bait (half of a Honey Loop) buried in each, were placed in the two smaller compartments, and the partitions were removed allowing rats to approach the bowls in turn, uncover and eat both of the cereal pieces. This was repeated for a total of six trials. If the rat did not uncover the rewards from both bowls within 10 min of being given access to them, then the partitions were lowered, both bowls were rebaited and the trial repeated. To ensure that the rats would respond promptly during sessions when escitalopram would be administered, they were given additional training in the test. In stage 2, rats were exposed to each of the exemplars that they would encounter the following day during testing. The exemplars were paired as they would be during testing, but with odours and media presented separately (see Table 1). Both bowls were baited with half a Honey Loop, and rats were exposed to each pair twice (sides switched). The rat was given 10 min to obtain the reward from each bowl as in stage 1 of the training. During stage 3 the rat learned two simple discriminations, in which

**Table 1** The list of exemplars used and their pairings. Exemplars are paired to reduce the complexity of counterbalancing. Medium pairings were chosen to minimise olfactory differences within a pair

The exemplars used

Dimension	Training pairing	Pairing 1	Pairing 2
Odour	O9—Mint O10—Oregano	O1—Cinnamon O2—Ginger	O3—Sage O4—Paprika
Medium	M9—Polystyrene M10—Confetti	M1—Coarse Tea M2—Fine Tea	M3—Sand M4—Grit

the bowls had different odours (the sawdust was scented with mint or oregano) or were filled with different digging media (paper confetti or small polystyrene pieces), and the rat had to learn which of the two bowls was baited.

The side of the baited bowl was determined pseudo-randomly for each trial, with a constraint being that there were no more than three consecutive trials with the reward on the same side. If the rat dug in the correct bowl, the latency to dig was recorded and that trial was recorded as correct. The trial terminated when the rat returned to the waiting area of the box, at which point the barrier was lowered and the bowls re-baited. If the rat dug in the incorrect bowl, the latency to dig was recorded and the trial was marked as incorrect, but the rat was still permitted to continue to explore that bowl; the trial was only terminated when the rat returned to the waiting area, at which point the barrier was lowered. For the initial four trials at each stage of the test, the rat was allowed dig in the correct bowl to recover the reward after an initial incorrect response; after the fourth trial an incorrect response terminated the trial. Whether the rat initiated digging in the first bowl encountered or whether he explored both bowls prior to initiating digging was also recorded. The rat was given up to 10 min to uncover the reward from the baited bowl; if the reward was not uncovered the partitions were lowered and the experimenter waited until the rat showed interest again.

Criterion performance was six consecutive correct trials (the probability of making a correct choice 6 times consecutively by chance is 0.015), which could include the first four trials.

#### 2.4.2 Behavioural Testing

On the first test day, the rat performed two series of three discriminations (Table 2). Both series consisted of a compound discrimination (acquisition (ACQ)), in which the rat must learn a novel discrimination between two exemplars of one dimension, ignoring the exemplars of an irrelevant dimension; a reversal (novel-reversal

**Table 2** Within the two series of discriminations, the order of exemplar pair exposure and whether odour or medium was rewarded in series 1 or 2 was counterbalanced, and matched between groups

An example of the order of exemplar pair exposure		
Discrimination	Relevant dimension exemplars	Irrelevant dimension exemplars
Series 1 Acquisition (ACQ)	<b>O1/O2</b>	M1/M2
Reversal (REV)	<b>O2/O1</b>	M1/M2
Reversal (BACK)	<b>O1/O2</b>	M1/M2
Series 2 Acquisition (ACQ)	<b>M3/M4</b>	O3/O4
Reversal (REV)	<b>M4/M3</b>	O3/O4
Reversal (BACK)	<b>M3/M4</b>	O3/O4

(REV)), where the exemplars remain the same as in the ACQ, but the correct and incorrect exemplars are reversed; a second reversal (reversal-back (BACK)), where the correct/incorrect status of the exemplars is reversed such that the discrimination is the same as during the ACQ stage. In the second series of three discriminations, novel stimuli were used, and the dimensional relevance to solving the discriminations was swapped.

The task advanced to the next stage when the rat had reached criterion (six correct trials consecutively). The procedure followed was the same for each stage: for the first four trials, the rat had the opportunity to dig in the correct bowl if it had first dug in the incorrect bowl. Thereafter, when the rat started to dig in either bowl, the partition to the other compartment was lowered to prevent access to the other bowl. The trial was not terminated until the rat returned to the waiting area. If the rat did not dig within 10 min, the partitions were lowered, separating the rat from the bowls. The trial was aborted and recorded as ‘non dig’.

Subsequent testing followed the same protocol, although rats did not need to be trained again for these tests.

#### **2.4.3 Counterbalancing**

Order of exposure to the dimensions (i.e., initial rewarded dimension being odour or medium) and to the exemplars was not fully counter-balanced due to the number of exemplars and their possible combinations. Exemplars were presented in pre-assigned pairs (see Table 1) and within each dose, starting dimension and order of presentation of pairs was balanced. Counterbalancing was matched between lesioned and control rats.

#### **2.4.4 Drug Administration**

Rats were administered a 1 ml/kg (s.c.) injection of sterile saline on the two days prior to the first test. On the day of testing, rats were administered either a 1 ml/kg (s.c.) injection of sterile saline or a 1, 2, or 4 mg/kg (s.c.) injection of escitalopram (in sterile saline at 1 ml/kg) 30 min prior to testing. Administration of dose was counterbalanced according to a Latin square design. Each rat received each dose once, with the control and lesioned groups matched.

#### **2.4.5 Histology**

Rats were transcardially perfused with 4% paraformaldehyde in 0.1 M phosphate buffer (PB) after anaesthesia with 0.8 ml Dolethal. The brains were sectioned (50 µm) and stained for neuronal nuclei (NeuN) and counterstained with cresyl violet to map lesion extent, following standard protocols reported previously.

### **2.4.6 Data Analysis**

Trials to criterion data (excluding non-digs) were analysed by repeated measures ANOVA (SPSS v 19.0) with dose (4 levels: vehicle, 1, 2 and 4 mg/kg escitalopram), discrimination series (2 levels: first and second) and stage (3 levels: ACQ, REV and BACK) as within subject variables, and group (2 levels: control and lesion) as between subjects variable.

## **2.5 *Experiment 2: Fos Activity After 1 mg/kg Escitalopram***

### **2.5.1 Behavioural Training**

Between 10 and 30 days after surgery, eight rats (lesion, n = 4; control, n = 4) were trained and tested on the reversal learning task. A further eight rats (lesion, n = 4; control, n = 4) were designated as their yoked controls. As rats were pair-housed, within each pair, one rat was designated to perform the reversal learning task, and the other would be its yoked control. The pair were trained and tested simultaneously. The eight rats that underwent the reversal learning task were trained and tested as described in experiment 1. The eight yoked controls underwent stage 1 of training as previously described, but thereafter training was altered. For stage 2 of training, yoked control rats dug in and obtained a single reward from each of two identical sawdust-filled bowls, an equal number of times to the reversal learning rat. During stage 3 of training, the yoked control rat was given access to two identical sawdust-filled bowls, each containing reward. Each time the reversal learning rat obtained reward, the yoked control rat was granted access to both bowls to obtain reward from one of them.

### **2.5.2 Behavioural Testing**

The day after training, the reversal learning rats performed the two series of three discriminations as described in experiment 1. For the duration of testing, whenever the reversal learning rat obtained a reward the yoked control rat was given access to two identical sawdust-filled bowls and allowed to obtain reward from one of them.

### **2.5.3 Counterbalancing**

With only two reversal learning rats in each condition counterbalancing of exemplars was not possible. Therefore, exemplars were presented in pre-assigned pairs as in experiment 1 and the order of exposure for all rats was the same.

### 2.5.4 Drug Administration

Rats were administered a 1 ml/kg (s.c.) injection of sterile saline for two days prior to testing. On the day of testing, rats were administered either a 1 ml/kg (s.c.) injection of sterile saline or a 1 mg/kg (s.c.) injection of escitalopram (1 mg/ml in sterile saline) 30 min prior to testing. There were therefore four conditions with two reversal learning rats and two yoked controls in each: control/saline; control/escitalopram; OFC lesion/saline; and OFC lesion/escitalopram.

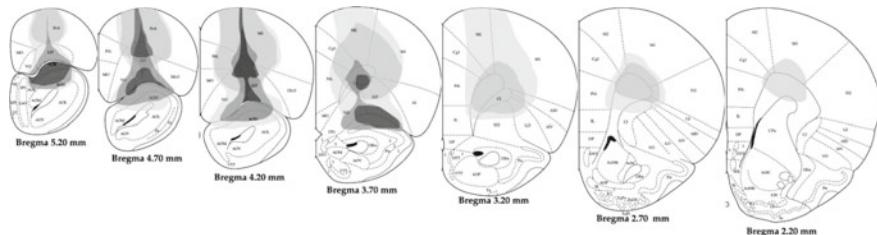
### 2.5.5 Histology

Rats were transcardially perfused 90 min after completion of testing and brain sections stained for neuronal nuclei (NeuN) and counterstained with cresyl violet as for experiment 1. For Fos immunoreactivity, sections were treated initially as for NeuN, except they were incubated in goat anti-Fos (dilution 1:8000) on a stirrer for 1 night, followed, after a 5 min wash in sterile PBS, by incubation on a shaker for one hour in rabbit anti-goat biotinylated secondary antibody (vector IgG solution at 5 µl/ml ADS). After washing in 0.1 M PBS again, sections were incubated on a stirrer in Vectastain ABC complex (as above) for a further hour. Sections were then washed in 0.1 M PBS again, and finally immersed in Sigma Fast DAB tablets for approximately 10 min, with the time being determined by visual inspection of the tissue. The tissue was removed when background staining was light but neurons were clearly visible. Sections were washed again in 0.1 M PBS and then mounted on treated glass slides, air-dried and cover-slipped with DPX. Fos positive neurons in the prelimbic area of the medial prefrontal cortex (mPFC) and in the OFC were counted by H. Lundbeck A/S.

### 2.5.6 Data Analysis

Trials to criterion data were analysed by repeated measures ANOVA (SPSS v 19.0) with stage (3 levels: ACQ, REV and BACK) as within subject variables, and dose (2 levels: vehicle and 1 mg/kg escitalopram) and group (2 levels: OFC lesion and control) as between subject variables. Discrimination series was not used as a within subject variable: whilst all rats completed the first series of discriminations, not all rats completed all stages in the second. A mean of the data collected over the two series was therefore used where rats had completed those stages.

Area-corrected Fos activation counts were analysed by repeated measures ANOVA with side (2 levels: right and left) as the within-subjects variable, and dose (as above), group (as above) and behaviour (2 levels: reversal learning and yoked control) as between-subjects variables.



**Fig. 1** Coronal schematics of the rat brain (adapted from Paxinos and Watson 2006) showing greatest extent of (light grey), typical (mid grey) and smallest (dark grey) lesion damage for rats from experiment 1

### 3 Results

### 3.1 Experiment 1

### **3.1.1 Histology**

Lesion placement was visualised in the NeuN/cresyl violet stained sections (Fig. 1). Approximately half of the lesions were positioned more dorsally, with the other half positioned ventrally. All lesioned rats showed cell loss in ventral and lateral OFC regions from bregma +5.00 to +3.50.

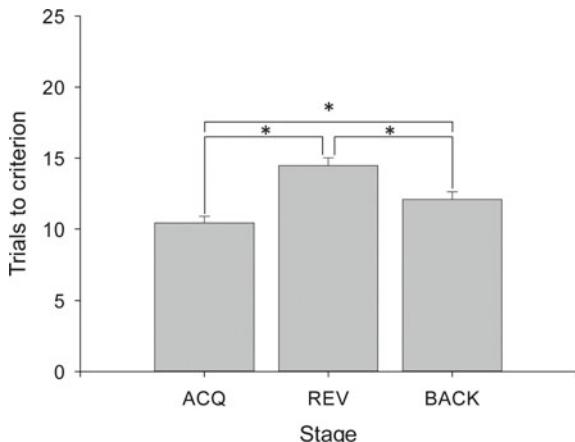
### 3.1.2 Behavioural Testing

Within a test, rats performed both discrimination series equally—there was no main effect of discrimination series ( $F_{(1,9)} = 0.8$ , not significant (ns)), nor was there any interaction between discrimination series and any other variable. Data are therefore presented collapsed across discrimination series. There was a main effect of stage ( $F_{(2,18)} = 29.6, p < 0.05$ ) and contrasts confirmed that new acquisitions required fewer trials to criterion than either novel-reversal ( $F_{(1,9)} = 46.7, p < 0.05$ ) or reversal-back ( $F_{(1,9)} = 18.0, p < 0.05$ ). In addition, reversal-back was learned more readily than novel reversals ( $F_{(1,9)} = 16.8, p < 0.05$ ) (Fig. 2).

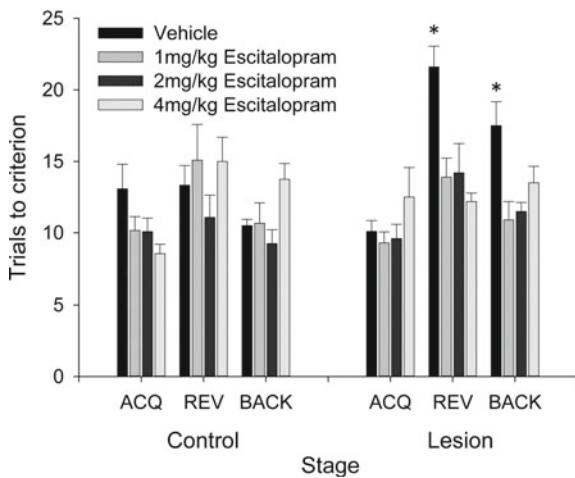
There was a three-way interaction between dose, group and stage ( $F_{(6,54)} = 4.9, p < 0.05$ ) (Fig. 3) in the context of no significant main effect of group ( $F_{(1,9)} = 3.8, \text{ns}$ ) or interactions of dose and group ( $F_{(3,27)} = 2.4, \text{ns}$ ), dose and stage ( $F_{(6,54)} = 1.3, \text{ns}$ ) or stage and group ( $F_{(2,18)} = 2.8, \text{ns}$ ). To probe this three-way interaction, corrected ANOVAs (using the error term from the omnibus ANOVA) were performed for each dose, with stage as within, and group as between-subjects variables.

In the vehicle condition, there was an interaction of stage and group ( $F_{2,54} = 5.9, p < 0.05$ ). Planned contrasts confirmed what is clear from Fig. 3: there was a difference

**Fig. 2** Mean + SEM trials to criterion from experiment 1 collapsed across group and dose to show the main effect of stage. Reversal stages required more trials to criterion than the acquisition (ACQ). The novel reversal stage (REV) also required more trials to criterion than the reversal back (BACK)

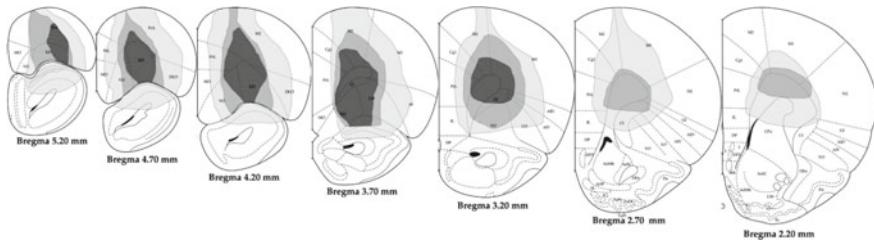


**Fig. 3** Mean + SEM trials to criterion from experiment 1 collapsed across discrimination series. Lesioned rats were impaired relative to control rats at the REV and BACK stages only after vehicle administration. Escitalopram at all doses ameliorate the effects of the lesion without affecting control rat performance (Paxinos and Watson 2006)



between the groups at the REV ( $F_{6,54} = 10.6, p < 0.05$ ) and BACK ( $F_{6,54} = 7.6, p < 0.05$ ) stages, but not in the ACQ stage ( $F_{6,54} = 1.4, \text{ns}$ ).

In the three escitalopram conditions, there were no main effects of group, nor any interactions between group and stage. OFC-lesioned rat reversal performance is only impaired relative to control rats in the vehicle group: escitalopram administration at all three doses ameliorates the effects of the OFC lesion on both novel-reversals and reversals-back.



**Fig. 4** Coronal schematics of the rat brain (adapted from Paxinos and Watson 2006) showing greatest extent of (light grey), typical (mid grey) and smallest (dark grey) lesion damage for rats from experiment 2

### 3.2 Experiment 2

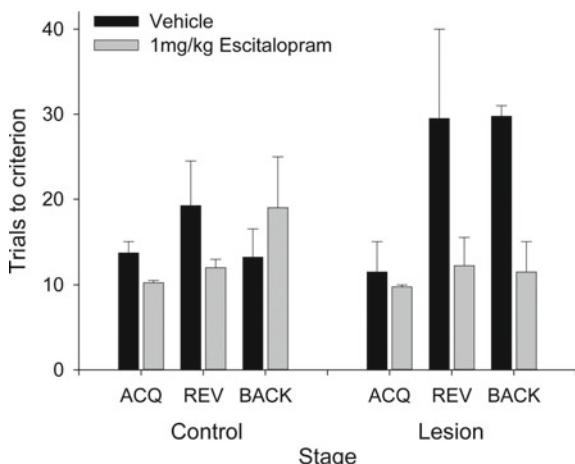
#### 3.2.1 Histology

Lesion placement was visualised in the NeuN/cresyl violet stained sections (Fig. 4). All lesioned rats showed cell loss in ventral and lateral OFC regions from bregma + 5.00 to +3.50.

#### 3.2.2 Behavioural Testing

Figure 5 shows the number of trials to criterion for each stage at each dose. All rats completed the first series of discriminations, but not all completed the second series within the 90-min testing window. Data were collapsed across discrimination series (acquisition, novel reversal (REV) and reversal back (BACK)) where possible. No

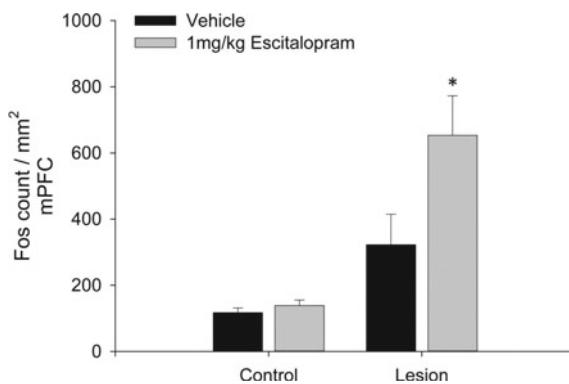
**Fig. 5** Mean + SEM trials to criterion from experiment 2 collapsed across discrimination series. The pattern of impairment is similar to that seen in experiment 1, with the same beneficial effect of escitalopram administration on reversal performance in the lesioned rats



statistically significant effects were found, likely due to variability within the small sample size, although the visual trend in the data suggests escitalopram is improving reversal learning in the lesioned rats as in experiment 1.

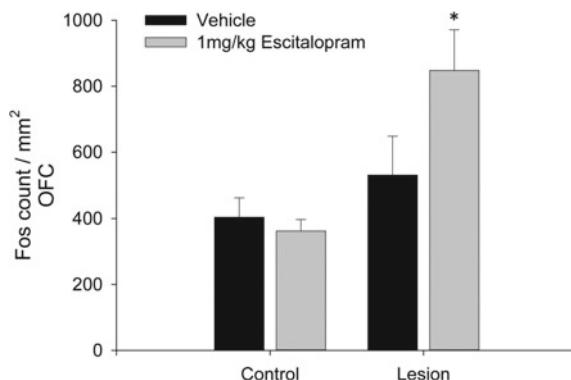
### 3.2.3 Fos Expression

Fos positive neurons were counted in the mPFC and OFC. Figure 6 shows area corrected (count/mm<sup>2</sup>) Fos counts for mPFC. There was an interaction between drug and group ( $F_{(1,8)} = 6.87, p < 0.05$ ): OFC-lesioned rats show greater Fos expression in mPFC than controls and escitalopram induces a further increase in Fos expression in rats with OFC lesions. The same pattern was also seen in the OFC (see Fig. 7):



**Fig. 6** Mean + SEM Fos count/mm<sup>2</sup> in the mPFC collapsed across side (behaving and yoked rats combined). More Fos activity was recorded in the lesioned rats' mPFC regardless of behaviour. Escitalopram increased Fos activity in the lesioned rats (regardless of whether they were performing a task or yoked control—not shown) without effect in the control rats (\* interaction of group and dose,  $p < 0.05$ )

**Fig. 7** Mean + SEM Fos count/mm<sup>2</sup> in the OFC collapsed across side (behaving and yoked rats combined). More Fos activity was recorded in the lesioned rats' mPFC regardless of behaviour. Escitalopram increased Fos activity in the lesioned rats regardless of behaviour, without effect in the control rats



an interaction between group and dose ( $F_{(1,8)} = 5.75, p < 0.05$ ) arose because OFC-lesioned rats show greater Fos expression in surviving areas of OFC than was seen in the intact OFC of controls. Escitalopram induces a further increase in activation of remaining OFC neurons in OFC-lesioned rats.

## 4 Discussion

The aim of this study was to examine the nature of cognitive structures in the rat, looking specifically at the underlying processes and cognitive structures in reversal learning. As reported previously (Chase et al. 2012; McAlonan and Brown 2003; Tait and Brown 2007; Tait et al. 2018), rats with non-selective OFC lesions are impaired relative to controls during compound discrimination reversal learning. Our new data demonstrates that this impairment occurs equally in both novel reversals and reversals returning to a previously learned discrimination. This impairment is ameliorated by administration of the SSRI, escitalopram, at all doses investigated (1, 2 and 4 mg/kg).

Expression of Fos protein in both the mPFC and intact areas of OFC was increased in rats with OFC lesions. Escitalopram at 1 mg/kg potentiated this lesion-induced Fos increase, regardless of the behaviours investigated, but had no effect on Fos expression in control rats.

### 4.1 Reversal Learning

Previous investigations of serial reversal learning in rodents have involved consecutive stages requiring alternation of responding, typically requiring a spatial discrimination (e.g., Béracochea et al. 2003; Boulougouris et al. 2007; Stalnaker et al. 2007). Serial discrimination reversal learning using visual stimuli has been reported in primates (e.g., Clarke et al. 2007) and using olfactory stimuli in rats (Kinoshita et al. 2008; Schoenbaum et al. 2003). In these studies, stimuli were “simple”, in that there was one correct and one incorrect with no deliberately embedded irrelevant information—i.e., any discriminable feature of a stimulus could be used to predict that stimulus’ reward status. Our task design adapted the rodent ID/ED attentional set-shifting task, and therefore used compound stimuli—i.e., there was a dual dimensionality to the stimuli, with one dimension’s features predicting reward status and the other being uncorrelated with reward status. A compound discrimination reversal must be more difficult than a simple discrimination reversal due to the additional requirement to filter out irrelevant information. Impaired performance at these reversal stages can therefore reflect a reduced ability to either adapt to changes in stimulus reward status, or to filter out this irrelevant information.

In a typical serial reversal learning task, there are several consecutive reversals, with the subject required to switch and back and forth. Improvements occur with

successive reversals. As our task design included a novel discrimination between four reversal stages, the third reversal is similar to the first (both are novel-reversals), and the fourth reversal is similar to the second (both are reversals-back). That we observed no difference in performance between the first and second discrimination series reversals, but that there is a difference between novel-reversals and reversals-back, suggests that a learning set did not form. Our data thus demonstrate that novel-reversals require more trials to learn than reversals-back. This difference likely arises from the reversals-back being facilitated by familiarity with the particular stimuli, rather than learning about reversals (which would also have benefitted the subsequent reversals).

## ***4.2 The Effects of OFC Lesions on Reversal Learning***

The role of the OFC in reversal learning in rats is well documented (Ghods-Sharifi et al. 2008; Kim and Ragozzino 2005; McAlonan and Brown 2003; Schoenbaum et al. 2002, 2003; Murray et al. 2007; Chase et al. 2012; Tait and Brown 2007). The processes underlying OFC lesion-induced reversal learning impairments are less clear. We have previously reported that OFC lesions impair reversal learning in compound discrimination reversal learning (McAlonan and Brown 2003) during a test of attentional set-shifting, and that this impairment likely does not arise from perseverative responding to previously rewarded stimuli (Tait and Brown 2007). However, rats with OFC lesions do not benefit from forming an attentional set—there was no difference in performance between intradimensional (ID) and extradimensional (ED) shift stages in the OFC-lesioned rat (McAlonan and Brown 2003; Chase et al. 2012). We have further reported that excitotoxic lesions of the nucleus basalis magnocellularis of the basal forebrain also impair reversal learning and also result in no difference between ID and ED shift performance (Tait and Brown 2008). In these lesion studies where the ID/ED differences are lost, there is no evidence of a difference between control and lesion group ED shift performances. Instead the data suggest that the ID/ED difference is lost because of worsening performance at the ID stage. Whilst the experimental design of these studies preclude drawing strong conclusions about set-formation, it would be predicted that rats that fail to form an attentional set would not show a shifting cost at the ED stage—i.e., rats try to solve the ID and ED shift stages with no a priori dimensional bias, and there should therefore be no difference in performance between those two stages. These data then imply one of two possibilities: either OFC lesions and/or basal forebrain lesions directly impair both reversal learning and attentional set-formation; or impairments in reversal learning induce impairments in attentional set-formation. To partially answer this question, we reported that OFC lesions do impair set-formation in rats independently of reversal learning in a variant of the ID/ED task with multiple ID stages and no reversal stages (Chase et al. 2012). We cannot yet, however, rule out the reverse: the possibility that impairments in set-formation result in a reduced reversal learning ability. However, given that there are considerable data demonstrating OFC

lesion-induced reversal learning deficits outwith tests of compound discrimination reversal learning, we are confident to conclude that the OFC-lesion induced deficits in reversal learning that we report here are a reflection of a fundamental impairment in reversal learning. That OFC-lesioned rats may find compound discrimination reversal learning more difficult than simple discrimination reversal learning because of an additional reduced ability to disregard the irrelevant information present in a compound discrimination is a possibility, but unlikely to be the sole source of the impairment. Furthermore, whilst our task is based on a modified version of the rodent ID/ED task, it does not contain measures of attentional set-formation or set-shifting per se, so attempts to draw conclusions on such would be overly speculative.

### 4.3 *The Effects of Escitalopram on Reversal Learning*

Increasing the availability of serotonin improves reversal learning in OFC-lesioned rats, and does so in both novel-reversal and reversals-back. Whilst there is a consensus that serotonergic (5-HT) manipulations impact reversal learning, reported results depend not just on the specific manipulation, but also on the form of reversal learning tested. Tryptophan depletion does not impair spatial reversal learning in rats (van der Passe and Feenstra 2008), but inhibition of tryptophan hydroxylase by *para*-chlorophenylalanine does impair compound discrimination reversal learning in an attentional set-shifting task (Lapiz-Bluhm et al. 2009). In primates, 5,7-dihydroxytryptamine lesions of OFC impairs visual discrimination reversal learning—both in simple discrimination serial reversal learning and compound discrimination reversal learning during an attentional set-shifting task (Clarke et al. 2007). Increasing endogenous 5-HT improves reversal learning in rodents: citalopram, consisting of both the r- and s-citalopram enantiomers, improves probabilistic reversal learning after both acute and sub-chronic dosing regimes (Bari et al. 2010). Whilst an acute administration of 1 mg/kg citalopram impairs, a higher dose (10 mg/kg) improves, probabilistic reversal learning performance. Lower doses of escitalopram, being more potent than citalopram, would be expected to produce similar effects to higher doses of citalopram. Hence, the fact that we report amelioration of OFC lesion-induced reversal learning impairments at an escitalopram dose of 1 mg/kg should not be considered a conflict with the data that show that the same dose of citalopram impairs reversal learning. Indeed, Bari et al. (2010) discuss evidence that low levels of citalopram induce different outcomes on PFC 5-HT availability, which may explain their reported impairment. It has also been reported that vortioxetine, a SSRI and serotonin receptor modulator, ameliorates reversal learning in an attentional set-shifting task in rats subjected to freezing stress (Wallace et al. 2014).

Reversal learning was thought to involve two distinct phases (see Sutherland and Mackintosh 1971): initially, after the change in the reinforcement contingency is detected, the response must extinguish; subsequent to a period of responding randomly, the new association is gradually learned. We recently demonstrated that this is overly simplistic: responding ‘at chance’ while seeking a solution is unlikely

to be governed by responding ‘by chance’ (Dhawan et al. 2019). While reversal learning paradigms can depend on model-free learning, they may also involve model-based processes (Doll et al. 2012; Izquierdo et al. 2017; Dhawan et al. 2019). In serial reversal learning tasks, performance improves with each reversal, as if the animal learns, over-and-above the particular S+/S– attribute, a win-stay/lose-shift rule, which Harlow (1949) referred to as a ‘learning set’. In the present study, the rats performed a reversal and then reversed back only once, but already there was a learning benefit. However, it is unlikely that this benefit arose from learning a ‘win-stay rule’ because it did not extrapolate to either the first reversal of a subsequent novel discrimination or the reversal back of that second discrimination reversal.

That neither OFC lesions, nor administration of escitalopram, affects the relationship between novel-reversals and reversals-back implies that there are similar processes involved in each form of reversal—or, more specifically, processes that are affected by OFC lesions and interactions with escitalopram mediate both reversing and reversing back—and whilst the task is sensitive enough to distinguish between novel-reversals and reversals-back, it is not sensitive enough to elucidate differences after OFC lesions and escitalopram administration.

#### **4.4 Fos Activity**

The data from Fos expression suggest that there is increased, behaviourally independent, activation in both mPFC and OFC after OFC lesions, and that this increased activity is augmented by escitalopram with no significant effect on control animals. The Fos expression reported here is similar in pattern to that seen in surviving mPFC neurons after administration of the atypical antipsychotic, asenapine (Tait et al. 2009), to rats with mPFC lesions. Specifically, rats with mPFC lesions show increased activity in surviving mPFC neurons—an effect that is augmented by administration of asenapine—but that is again behaviourally independent. The similarity of the activation pattern may suggest that both drugs act through overlapping mechanisms on the mPFC, i.e. escitalopram by increasing serotonin levels and asenapine by modulating activity of serotonin receptors (Homberg 2012).

The increased mPFC and OFC Fos expression in the rats with OFC lesions was seen both when they were performing discrimination learning and reversals and also in yoked controls. Consequently, we can conclude that this expression is not a marker of activity driven by the cognitive processes underlying discrimination and reversal learning. It is likely then that there is increased recruitment of PFC neurons resulting from the lesion irrespective of the cognitive demands on the rats.

In intact rats, there was similarly no difference in Fos expression in rats performing the task or their yoked controls. This suggests that the cognitive processes mediated by these brain regions likely require low levels of activity from a relatively large pool of available neurons. Thus, our observations of low levels of Fos expression in the control rats arise because few neurons are activated to a sufficient threshold that Fos is expressed to a detectable level. In lesioned rats, with fewer PFC neurons, there

must be increased recruitment of surviving neurons in order for cognition to approach normal levels—more neurons need to activate to the threshold level where detectable Fos is expressed because there are fewer neurons to fulfil their respective roles. In the case of OFC-lesioned rats, this increased expression in a reduced number of neurons reflects increased neuronal activity that is insufficient to normalise reversal learning. However, escitalopram facilitates even greater PFC activity than could occur otherwise, and this increased activity is sufficient to normalise reversal learning in the OFC-lesioned rats. That we observed increased Fos activity in the mPFC of the OFC-lesioned rats, as well as the OFC, is a reminder that a network of brain regions underlies complex cognition and behavioural flexibility. mPFC neurons may be recruited to compensate for the functions that are impaired when the OFC is damaged. The mPFC, being adjacent to the OFC, was also damaged to some extent in most of the lesioned rats. Although this incidental mPFC damage did not result in the same behavioural profile associated with targeted mPFC, it is possible that this is due to compensatory elevation of mPFC activity, as indicated by increased Fos activation, in the surviving mPFC neurons. In both the case of asenapine-treated mPFC-lesioned rats and escitalopram-treated OFC-lesioned rats, behaviourally independent drug-induced increases in activity in surviving neuronal populations likely facilitate the cognitive processes that have been impaired by damage, but do not reflect activity actually driven by the undertaking of those cognitive processes.

The fact that reversal learning can be readily measured in different species, using species appropriate stimuli and responses, makes it a particularly valuable test for translational psychopharmacological research (see Izquierdo et al. 2017). Serial reversal learning is commonly used in non-human animals, often because this is a way to gather ‘additional data’ without recourse to lengthy training of new discriminations or the requirement to generate a large number of novel stimuli for testing. However, serial reversals should be thought of as more complex than simply repetition of the same thing. Reversing-back benefits from the additional familiarity with the stimuli, which is also seen if an animal is given additional post-criterion trials of overtraining. This effect is seen even in the absence of a benefit from the formation of ‘learning set’ (i.e., incorporating into the cognitive structure the concept that ‘reversals can occur’). We report here no evidence of a learning set following a single reversal/reversed back: subsequent reversals of new stimuli were not more rapidly acquired, even while reversing back was consistently more rapid than initial reversing. That notwithstanding, we conclude that reversal learning in OFC-lesioned rats is both an easily administered and sensitive test that can detect effects of serotonergic modulation on cognitive structures that are involved in behavioural flexibility.

**Acknowledgements** This study was funded by H. Lundbeck A/S.

## References

- Bari, A., Theobald, D. E., Caprioli, D., Mar, A. C., Aidoo-Micah, A., Dalley, J. W., et al. (2010). Serotonin modulates sensitivity to reward and negative feedback in a probabilistic reversal learning task in rats. *Neuropsychopharmacology*, 35(6), 1290–1301. <https://doi.org/10.1038/npp.2009.233>.
- Béracochéa, D., Celerier, A., Peres, M., & Pierard, C. (2003). Enhancement of learning processes following an acute modafinil injection in mice. *Pharmacology, Biochemistry and Behavior*, 76(3–4), 473–479. <https://doi.org/10.1016/j.pbb.2003.09.007>.
- Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology*, 39, 15–22.
- Birrell, J. M., & Brown, V. J. (2000). Medial frontal cortex mediates perceptual attentional set shifting in the rat. *Journal of Neuroscience*, 20(11), 4320–4324.
- Boulougouris, V., Dalley, J. W., & Robbins, T. W. (2007). Effects of orbitofrontal, infralimbic and prelimbic cortical lesions on serial spatial reversal learning in the rat. *Behavioural Brain Research*, 179(2), 219–228. <https://doi.org/10.1016/j.bbr.2007.02.005>.
- Brigman, J. L., Mathur, P., Harvey-White, J., Izquierdo, A., Saksida, L. M., Bussey, T. J., et al. (2010). Pharmacological or genetic inactivation of the serotonin transporter improves reversal learning in mice. *Cerebral Cortex*, 20(8), 1955–1963. <https://doi.org/10.1093/cercor/bhp266>.
- Chase, E., Alexander, Tait, D. S., & Brown, V. J. (2012). Lesions of the orbital prefrontal cortex impair the formation of attentional set in rats. *European Journal of Neuroscience*, 36(3), 2368–2375. <https://doi.org/10.1111/j.1460-9568.2012.08141.x>.
- Clarke, H. F., Walker, S. C., Dalley, J. W., Robbins, T. W., & Roberts, A. C. (2007). Cognitive inflexibility after prefrontal serotonin depletion is behaviorally and neurochemically specific. *Cerebral Cortex*, 17(1), 18–27. <https://doi.org/10.1093/cercor/bjh120>.
- Dhawan, S. S., Tait, D. S., & Brown, V. J. (2019). More rapid reversal learning following overtraining in the rat is evidence that behavioural and cognitive flexibility are dissociable. *Behavioural Brain Research*, 363(January), 45–52. <https://doi.org/10.1016/J.BBR.2019.01.055>.
- Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current Opinion in Neurobiology*, 22(6), 1075–1081. <https://doi.org/10.1016/j.conb.2012.08.003>.
- Floresco, S. B., & Jentsch, J. D. (2011). Pharmacological enhancement of memory and executive functioning in laboratory animals. *Neuropsychopharmacology*, 36(1), 227–250. <https://doi.org/10.1038/npp.2010.158>.
- Ghods-Sharifi, S., Haluk, D. M., & Floresco, S. B. (2008). Differential effects of inactivation of the orbitofrontal cortex on strategy set-shifting and reversal learning. *Neurobiology of Learning and Memory*, 89(4), 567–573. <https://doi.org/10.1016/j.nlm.2007.10.007>.
- Gilmour, G., Arguello, A., Bari, A., Brown, V. J., Carter, C., Floresco, S. B., et al. (2013). Measuring the construct of executive control in schizophrenia: Defining and validating translational animal paradigms for discovery research. *Neuroscience and Biobehavioral Reviews*, 37(9), 2125–2140. <https://doi.org/10.1016/j.neubiorev.2012.04.006>.
- Harlow, H. F. (1949). The formation of learning sets. *The Psychological Review*, 56(1), 51–65. <https://doi.org/10.1037/h0062474>.
- Homberg, J. R. (2012). Serotonin and decision making processes. *Neuroscience and Biobehavioral Reviews*, 36(1), 218–236. <https://doi.org/10.1016/j.neubiorev.2011.06.001>.
- Insel, T. R., Sahakian, B. J., Voon, V., Nye, J., Brown, V. J., Altevogt, B. M., et al. (2012). Drug research: A plan for mental illness. *Nature*, 483(7389), 269–269. <https://doi.org/10.1038/483269a>.
- Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., & Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience*, 345, 12–26. <https://doi.org/10.1016/j.neuroscience.2016.03.021>.
- Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology*, 89, 1–81.

- Kehagia, A. A., Murray, G. K., & Robbins, T. W. (2010). Learning and cognitive flexibility: Frontostriatal function and monoaminergic modulation. *Current Opinion in Neurobiology*, 20(2), 169–192. <https://doi.org/10.1016/j.conb.2010.01.007>.
- Kim, J., & Ragozzino, M. E. (2005). The involvement of the orbitofrontal cortex in learning under changing task contingencies. *Neurobiology of Learning and Memory*, 83(2), 125–133. <https://doi.org/10.1016/j.nlm.2004.10.003>.
- Kinoshita, S., Yokoyama, C., Masaki, D., Yamashita, T., Tsuchida, H., Nakatomi, Y., et al. (2008). Effects of rat medial prefrontal cortex lesions on olfactory serial reversal and delayed alternation tasks. *Neuroscience Research*, 60(2), 213–218. <https://doi.org/10.1016/j.neures.2007.10.012>.
- Lapiz-Bluhm, M. D. S., Soto-Piña, A. E., Hensler, J. G., Morilak, D. A. (2009). Chronic intermittent cold stress and serotonin depletion induce deficits of reversal learning in an attentional set-shifting test in rats. *Psychopharmacology*, 202(1–3), 329–341. <https://doi.org/10.1007/s00213-008-1224-6>.
- Lawrence, D. H. (1949). Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology*, 39(6), 770–784. <https://doi.org/10.1037/h0058097>.
- McAlonan, K., & Brown, V. J. (2003). Orbital prefrontal cortex mediates reversal learning and not attentional set shifting in the rat. *Behavioural Brain Research*, 146(1–2). <https://doi.org/10.1016/j.bbr.2003.09.019>.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24(1), 167–202. <https://doi.org/10.1146/annurev.neuro.24.1.167>.
- Murphy, F. C., Michael, A., & Sahakian, B. J. (2012). Emotion modulates cognitive flexibility in patients with major depression. *Psychological Medicine*, 42(7), 1373–1382. <https://doi.org/10.1017/S0033291711002418>.
- Murray, E. A., O'Doherty, J. P., & Schoenbaum, G. (2007). What we know and do not know about the functions of the orbitofrontal cortex after 20 years of cross-species studies. *Journal of Neuroscience*, 27(31), 8166–8169. <https://doi.org/10.1523/JNEUROSCI.1556-07.2007>.
- Murray, G. K., Cheng, F., Clark, L., Barnett, J. H., Blackwell, A. D., Fletcher, P. C., et al. (2008). Reinforcement and reversal learning in first-episode psychosis. *Schizophrenia Bulletin*, 34(5), 848–855. <https://doi.org/10.1093/schbul/sbm078>.
- Paxinos, G., & Watson, C. (2006). *The rat brain in stereotaxic coordinates*. 6th ed. Academic Press.
- Ravizza, S. M., & Carter, C. S. (2008). Shifting set about task switching: behavioral and neural evidence for distinct forms of cognitive flexibility. *Neuropsychologia*, 46(12), 2924–2935. <https://doi.org/10.1016/j.neuropsychologia.2008.06.006>.
- Schoenbaum, G., Nugent, S. L., Saddoris, M. P., & Setlow, B. (2002). Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. *NeuroReport*, 13(6), 885–890. <https://doi.org/10.1097/00001756-200205070-00030>.
- Schoenbaum, G., Setlow, B., Saddoris, M. P., & Gallagher, M. (2003). Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron*, 39(5), 855–867. [https://doi.org/10.1016/S0896-6273\(03\)00474-4](https://doi.org/10.1016/S0896-6273(03)00474-4).
- Stalnaker, T. A., Franz, T. M., Singh, T., & Schoenbaum, G. (2007). Basolateral amygdala lesions abolish orbitofrontal-dependent reversal impairments. *Neuron*, 54(1), 51–58. <https://doi.org/10.1016/j.neuron.2007.02.014>.
- Sutherland, N. Stuart, & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning*. New York: Academic Press Inc.
- Tait, D. S., & Brown, V. J. (2008). Lesions of the basal forebrain impair reversal learning but not shifting of attentional set in rats. *Behavioural Brain Research*, 187(1), 100–108. <https://doi.org/10.1016/j.bbr.2007.08.035>.
- Tait, D. S., & Brown, V. J. (2007). Difficulty overcoming learned non-reward during reversal learning in rats with ibotenic acid lesions of orbital prefrontal cortex. *Annals of the New York Academy of Sciences*, 1121, 407–420. <https://doi.org/10.1196/annals.1401.010>.

- Tait, D. S., Marston, H. M., Shahid, M., & Brown, V. J. (2009). Asenapine restores cognitive flexibility in rats with medial prefrontal cortex lesions. *Psychopharmacology (Berl)*, 202(1–3), 295–306. <https://doi.org/10.1007/s00213-008-1364-8>.
- Tait, D. S., Bowman, E. M., Neuwirth, L. S., & Brown, V. J. (2018). Assessment of intradimensional/extradimensional attentional set-shifting in rats. *Neuroscience and Biobehavioral Reviews*, 89(February), 72–84. <https://doi.org/10.1016/j.neubiorev.2018.02.013>.
- van der Plasse, G., & Feenstra, M. (2008). Serial reversal learning and acute tryptophan depletion. *Behavioural Brain Research*, 186, 23–31. <https://doi.org/10.1016/j.bbr.2007.07.017>.
- Wallace, A., Pehrson, A. L., Sánchez, C., & Morilak, D. A. (2014). Vortioxetine restores reversal learning impaired by 5-HT depletion or chronic intermittent cold stress in rats. *International Journal of Neuropsychopharmacology*, 17(10), 1695–1706. <https://doi.org/10.1017/S1461145714000571>.
- Waltz, J. A. (2017). The neural underpinnings of cognitive flexibility and their disruption in psychotic illness. *Neuroscience*, 345, 203–217. <https://doi.org/10.1016/j.neuroscience.2016.06.005>.
- Whitehead, A. N. (1938). *Modes of thought*. The Macmillan Company.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Rat Ultrasonic Vocalizations as Social Reinforcers—Implications for a Multilevel Model of the Cognitive Representation of Action and Rats' Social World



Tobias Kalenscher, Lisa-Maria Schönfeld, Sebastian Löbner, Markus Wöhr, Mireille van Berkel, Maurice-Philipp Zech, and Marijn van Wingerden

**Abstract** Rats are social animals. For example, rats exhibit mutual-reward preferences, preferring choice alternatives that yield a reward to themselves as well as to a conspecific, over alternatives that yield a reward only to themselves. We have recently hypothesized that such mutual-reward preferences might be the result of reinforcing properties of ultrasonic vocalizations (USVs) emitted by the conspecifics. USVs in rats serve as situation-dependent socio-affective signals with important communicative functions. To test this possibility, here, we trained rats to enter one of two compartments in a T-maze setting. Entering either compartment yielded identical food rewards as well as playback of pre-recorded USVs either in the 50-kHz range, which we expected to be appetitive or therefore a potential positive reinforcer, or in the 22-kHz range predicted to be aversive and therefore a potential negative reinforcer. In three separate experimental conditions, rats chose between compartments yielding either 50-kHz USVs versus a non-ultrasonic control stimulus (condition 1), 22-kHz USVs versus a non-ultrasonic control stimulus (condition 2), or 50-kHz versus 22-kHz USVs (condition 3). Results show that rats exhibit a transient preference for the 50-kHz USV playback over non-ultrasonic control stimuli, as well as an initial avoidance of 22-kHz USV relative to non-ultrasonic control stimuli on trend-level. As rats progressed within session through trials, and across sessions,

---

T. Kalenscher (✉) · L.-M. Schönfeld · M. van Berkel · M.-P. Zech · M. van Wingerden  
Comparative Psychology, Institute of Experimental Psychology, Heinrich-Heine-Universität,  
40204 Düsseldorf, Germany  
e-mail: [Tobias.Kalenscher@hhu.de](mailto:Tobias.Kalenscher@hhu.de)

M. van Berkel · M. van Wingerden  
Social Rodent Lab, Institute of Experimental Psychology, Heinrich-Heine-Universität, 40204  
Düsseldorf, Germany

S. Löbner  
Institute of Linguistics and Information Science, Heinrich-Heine-Universität, 40204 Düsseldorf,  
Germany

M. Wöhr  
Behavioral Neuroscience, Experimental and Biological Psychology, Philipps-Universität, 35032  
Marburg, Germany

these preferences diminished, in line with previous findings. These results support our hypothesis that USVs have transiently motivating reinforcing properties, putatively acquired through association processes, but also highlight that these motivating properties are context-dependent and modulatory, and might not act as primary reinforcers when presented in isolation. We conclude this article with a second part on a multilevel cognitive theory of rats' action and action learning. The "cascade" approach assumes that rats' cognitive representations of action may be multilevel. A basic physical level of action may be invested with higher levels of action that integrate emotional, motivational, and social significance. Learning in an experiment consists in the cognitive formation of multilevel action representations. Social action and interaction in particular are proposed to be cognitively modeled as multilevel. Our results have implications for understanding the structure of social cognition, and social learning, in animals and humans.

**Keywords** Rats · Ultrasonic vocalization · Prosocial behavior · Reinforcement learning · Cognitive representation · Multilevel categorization · Cascades

## Part I: The Experiments

### 1 Introduction

Imagine you are passing through a heavy door that separates two parts of your university building. You notice that a person behind you also wants to walk through that heavy door. As an act of politeness, you hold the door open for him. Realizing this, he smiles at you and thanks you for your courtesy.

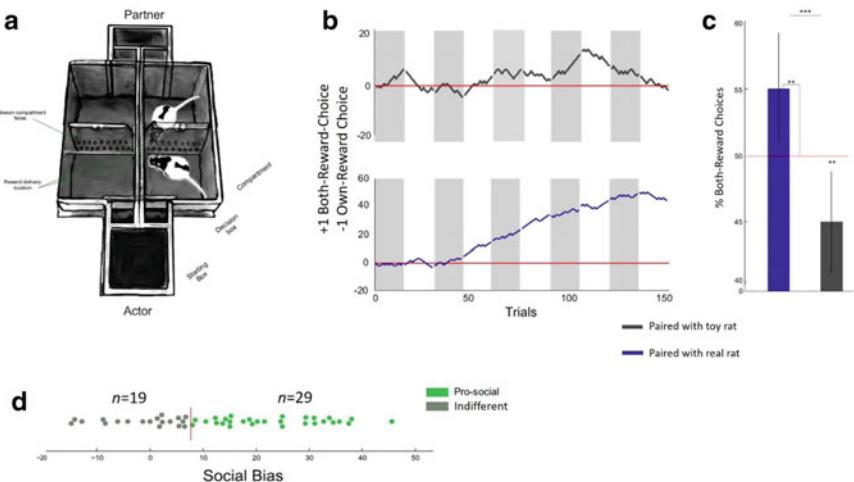
Why did you engage in such a (mildly) costly act of consideration? There are many putative reasons that may act in concert to support prosocial actions of this kind: adhering to the social norm that one should always help each other, following a generalized reciprocity principle as you may hope that someone else might hold a door open for you in the future, and working on your reputation as a friendly person. In addition, it is also possible that your behavior might be reinforced by the thankful response of the recipient of your help. According to this mechanism, you might have perceived the social signals emitted by him—his smile and his utterance of thankfulness—as rewarding, and, by consequence, the rewarding nature of these social signals might have increased the probability of repeating this helpful act in the future; that is, you will hold open the door for the next stranger again. This explanation is particularly intriguing as social signals are physical signals that can be multi-modally detected by the body's senses (smile: vision; words of thankfulness: auditory), yet they do not have primary hedonic value in themselves. Nevertheless, these signals have social significance that can influence, reinforce, and structure social behavior. In other words, stimuli like utterances and facial expressions can be

understood on different levels of conceptual meaning: physical, social and motivational salience that, jointly, is perceived as part of our social world, and thus govern social behavior.

It is intuitively evident that the ability to attach motivational and emotional significance to events in the social world is of prime importance for social cognition (Fiske & Taylor, 1984). However, our understanding of social cognition and its evolution is still incomplete. One likely reason is that we lack a proper conceptual framework to comprehend the cognitive, emotional and motivational processes associated with social stimuli. For instance, it is unclear how the attribution of motivational significance to physical stimuli is cognitively and neurally represented, and which features discriminate a social element from a non-social item. Simply speaking, individuals are influenced by social stimuli in a different way than by similar stimuli that lack social significance (e.g., a smile on the face of a display mannequin). However, it is unknown how individuals disambiguate between social and non-social stimuli, and how they attribute social and motivational significance to those stimuli.

Human social behavior is multi-faceted, is notoriously sensitive to cultural, experienced, cognitive, and gender-specific influences, and it is the outcome of a multitude of different motives. It is therefore imperative to avoid, or, at least, to control for possible confounding factors when studying social interaction. Although there is a rich literature on social cognition in the human domain (Fehr & Fischbacher, 2003; Fehr & Schmidt, 1999; Strombach et al., 2015), the best way to avoid confounding variations in cultural backgrounds, prior expectations and the tendency to show socially desirable behavior is to study social behavior in non-human animals. Moreover, we have recently argued in favor of complementing traditional human research with careful comparisons across species because such comparative approaches may offer answers to the question as to why humans make social and economic decisions as they do (Kalenscher & van Wingerden, 2011). Here, we plan to use rats as model organisms. Rats are highly social animals (Blanchard & Blanchard, 1990; Blanchard, Flannery & Blanchard, 1988) with a rich social behavior repertoire, including social play behavior (rough-and-tumble play; Siviy & Panksepp, 2011; Vanderschuren, Achterberg, & Trezza, 2016) and acoustic communication through ultrasonic vocalizations (USVs; Brudzynski, 2013; Wöhr & Schwarting, 2013). Furthermore, rats have been shown to exhibit prosocial behavior in various contexts and ways (Ben-Ami Bartal, Decety, & Mason, 2011; Hernandez-Lallement, van Wingerden, Marx, Srejic, & Kalenscher, 2015; Hernandez-Lallement, van Wingerden, Schäble & Kalenscher, 2016, 2017; Oberliessen et al., 2016; Rutte & Taborsky, 2007).

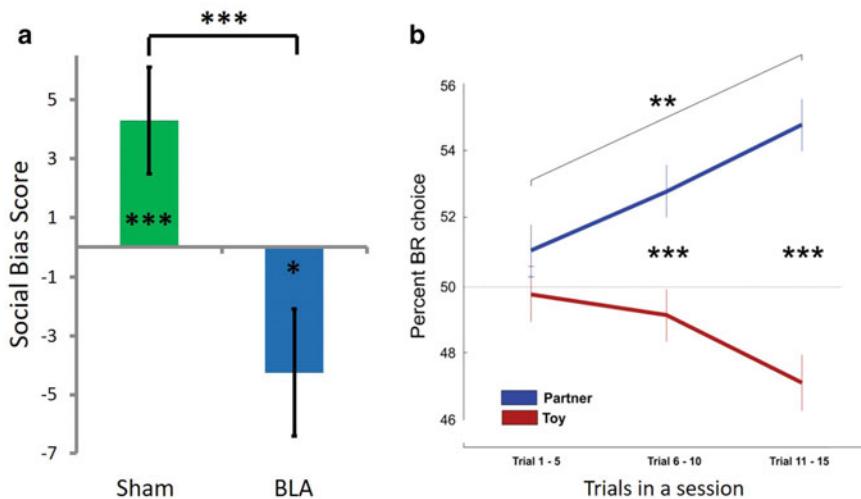
We have recently developed a prosocial choice task (PCT; Hernandez-Lallement et al., 2015) in which actor rats made non-costly decisions yielding a reward to a partner rat, or no reward to partner, respectively (Fig. 1a). Our results have shown that actor rats developed a preference for the both-reward option, yielding a reward for both the actor and the partner, over the own-reward option, yielding a reward only to the actor, but not the partner. Remarkably, this behavior was only displayed if the partner was a real rat, but not if it was a toy rat (Fig. 1b, c). The extent of prosocial behavior was not uniform across the animals; there was large individual variability between rats in their mutual-reward preference levels, as indicated by the



**Fig. 1** Prosocial choice task. **a** Double T-maze apparatus for quantifying mutual-reward preferences in pairs of rats. The actor rat chooses to enter either a both-reward compartment (both rats receive identical food rewards), or an own-reward compartment (only the actor receives a reward, but not the partner). The partner is always directed towards the opposite compartment facing the actor. Actor's and partner's compartments are separated by a transparent, perforated wall, allowing rats to see, hear and smell each other. **b** Example choice of one rat. The tally is increased by 1 every trial the actor rats makes a both-reward choice, and decreased by 1 every trial the actor rat makes an own-reward choice. Upper panel: actor rat paired with a toy rat. Lower panel: same actor rat paired with a real partner rat. **c** Mean percentage of both-reward choices, averaged across all rats and sessions. **d** Social bias scores. For each rat, the social bias score represents the percent differences in both-reward choices between the social and toy conditions. The social bias score can be interpreted as the added value of both-reward outcomes. The vertical bar represents the upper 95% confidence interval limit, which was based to categorize rats as prosocial (green dots; social bias scores exceeding the upper confidence interval limit), and indifferent (grey dots; social bias scores within the confidence interval limits) \*\*p < 0.05; \*\*\*p < 0.001. Adapted from Hernandez-Lallemand et al. (2015)

wide distribution of social bias scores (Fig. 1d; the social bias scores represent the percent differences in both-reward choices between the social and toy conditions and can be interpreted as the added value of both-reward outcomes).

In a follow-up lesion study, we found that mutual-reward preferences in rats disappeared after lesions of the basolateral amygdala (Hernandez-Lallemand et al., 2016), a brain structure implicated in emotional processes (LeDoux, 1994; LeDoux, Cicchetti, Xagoraris, & Romanski, 1990) as well as social and non-social reward representation (Chang et al., 2015; Janak & Tye, 2015). Our results showed that the social bias score, indicating the added value placed on mutual reward outcomes, turned negative in amygdala-lesioned rats (Fig. 2a) because they chose the both-reward option less often when paired with a real rat than when paired with a toy. This suggests that, in contrast to sham-lesioned animals, amygdala-lesioned rats



**Fig. 2** Social reinforcement learning in rats is amygdala-dependent. **a** Lesions to the basolateral amygdala (BLA) abolish mutual-reward preferences in rats, as indicated by negative social bias scores. Adapted from (Hernandez-Lallement et al., 2016). **b** Rats re-acquire both-reward (BR) preferences across trials in sessions after the compartment-contingency assignment was reversed. Adapted from (Hernandez-Lallement, van Wingerden, et al., 2017). \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$

failed to attach positive value to rewards delivered to partners; hence, the amygdala-lesioned animals behaved as if they had turned callous to the welfare of other rats (Hernandez-Lallement, van Wingerden, & Kalenscher, et al., 2017).

To better understand the emergence of mutual-reward preferences in non-lesioned control rats, we exploited the fact that the task contingencies were frequently reversed because the both-reward assignment to one of the two actor compartments was pseudo-randomized across testing days and rats (Hernandez-Lallement, van Wingerden, et al., 2017). We found that both-reward choices were at chance level in the first few trials after a contingency reversal, but gradually increased across trials (Fig. 2b). This finding suggests that rats re-learn which compartment yields reward to both rats after every contingency change. We hypothesized that such re-learning can be explained by standard reinforcement learning mechanisms (Sutton & Barto, 2012), with one notable exception. Because the payoff to the actor rat is always identical after own-reward or both-reward choices and in the partner- and the toy-conditions, and because the only difference between conditions is the social context, the reinforcer must be of social nature. Two non-mutually-exclusive mechanisms are conceivable by which social signals, whatever they are, may reinforce mutual-reward choices (Hernandez-Lallement, van Wingerden, et al., 2017): partner rats might emit social signals upon reward receipt that are rewarding to the actor rats, reinforcing the actor's behavior that yielded reward to the partner. In addition, missing out on reward might prompt the emission of distress or complaint signals by the partner that

are aversive to the actor rats, resulting in the avoidance of behaviors associated with these aversive complaint signals.

To date, it is unknown what kind of signals might serve as social reinforcers. However, several lines of evidence suggest that putative candidate signals for appetitive and aversive social reinforcement are rat USVs. Rats emit USVs in the 50-kHz range in positive affective states, for example, during rough-and-tumble play (Knutson, Burgdorf, & Panksepp, 1998; Lukas & Wöhr, 2015), tickling (Ishiyama & Brecht, 2016; Panksepp & Burgdorf, 2000), or after amphetamine injections (Burgdorf, Knutson, Panksepp, & Ikemoto, 2001; Engelhardt, Fuchs, Schwarting, & Wöhr, 2017). By contrast, rats vocalize in the 22-kHz range in negative affective states, e.g., during threatening situations or fear conditioning (Brudzynski & Ociepa, 1992; Calvino, Besson, Boehrer, & Depaulis, 1996; Parsana, Li, & Brown, 2012; Sales, 1972).

Rats show a strong, but short-lived orientation response and transient social approach behavior towards playback of pre-recorded 50-kHz USVs as well as avoidance of 22-kHz USV playback (Wöhr & Schwarting, 2007), and will perform more instrumental actions to obtain 50 kHz than 22 kHz USV playback (Burgdorf et al., 2008). Moreover, 50-kHz USV playback (Willuhn et al., 2014) or observing another rat getting rewarded (Kashtelyan, Lichtenberg, Chen, Cheer, & Roesch, 2014) elicits dopamine release in the nucleus accumbens, one of the key brain mechanisms for reinforcement learning (Parkinson, Robbins, & Everitt, 1996). In addition, 50- and 22-kHz signals elicit increases, or decreases respectively, in tonic firing activity in single neurons in the rat amygdala (Parsana et al., 2012), the very same brain structure whose integrity is necessary for expressing mutual-reward preferences in our PCT (Fig. 2a; Hernandez-Lallement, van Wingerden, & Kalenscher, 2017; Hernandez-Lallement et al., 2016).

Taken together, this evidence is in line with the hypothesis that 50- and 22-kHz USVs might serve as candidate signals for appetitive social reinforcement, or aversive social reinforcement respectively. Moreover, rats engaged in the PCT indeed vocalize, both in the 22 and 50 kHz domain (unpublished observations). We thus set out to investigate whether the playback of pre-recorded USVs in the context of the pro-social choice task setup would be as effective in driving choice behavior as the putative social signals emitted by partner rats in the full version of the PCT, while keeping task contingencies as close to the original PCT as possible. Specifically, we hypothesized that 50-kHz USV stimuli induce approach behavior and, thus, enhanced preference for outcomes associated with playback of 50-kHz USV playbacks. We furthermore hypothesized that 22-kHz USV stimuli are avoided by the rats, resulting in decreased preference for 22-kHz USV outcomes. In the following, we will present evidence that USVs, in contrast to similar acoustic stimuli of non-social nature, indeed have transient motivating properties and can drive spatial preferences linked to social outcomes as observed in the PCT.

Importantly, we go one step further than merely evaluating the social reinforcement hypothesis (Hernandez-Lallement, van Wingerden, et al., 2017). This hypothesis is useful in describing the cognitive mechanisms underlying mutual-reward preferences, but leaves open the question how rats cognitively construe a social situation

characterized by the presence of conspecifics and/or USVs. More specifically, it is unclear how a rat conceptually links and represents the several stimulus levels—the USV's physical dimension (rhythmic oscillations of air compression and deflation), their emotional level (the putative enjoyment or aversiveness of listening to USVs) and their motivational level (50-kHz USVs are wanted and prompt action to obtain them, 22-kHz USVs are avoided and prompt action to evade them)—into a coherent cognitive representation of a social situation. A promising approach to understand how rats cognitively construct their social world needs to transcend beyond the limitations of traditional reinforcement learning theory, and enter the realm of philosophy. Therefore, in addition to presenting evidence that rats attribute incentive value to USV playback, we will conclude this article with a theoretical perspective, inspired by linguistic theory, on the rat's cognitive representation of its social world. This theory addresses the point of multilevel cognitive representation of a social act, and how this can guide learning about social interaction.

## 2 Methods

### 2.1 Subjects

The experiment was approved by German authorities (Landesamt für Natur, Umwelt und Verbraucherschutz) and conducted according to the European Union Directive 2017/63/EU. Fifteen male Long-Evans rats (Charles River Laboratories, Calco, Italy) were housed in groups of three and kept under a reversed 12 h-dark/light cycle (lights off at 7 am). The housing room was at a constant temperature of  $20 \pm 2$  °C and a humidity of 60%. Rats received standard rodent laboratory food (Sniff, Soest, Germany), and water ad libitum. At the start of the experiment, food access was restricted to keep the animals at 90% of their free feeding body weight. Animals were randomly assigned to one of two groups differing in the stimulus material (see acoustic stimuli below): USV<sub>Type-1</sub> ( $n = 7$ ) and USV<sub>Type-2</sub> ( $n = 8$ ).

### 2.2 Experimental Setup

The playback experiment aimed to evaluate the effectiveness of playback of pre-recorded USVs in shaping spatial preferences as observed in the PCT (Hernandez-Lallement et al., 2015). As such, we employed the same behavioral setup as in the PCT, but with the following minor modifications. Each side of the maze (front: actor side; back: partner side) consisted of a start box measuring  $31 \times 20 \times 40$  cm leading via two doors to separate choice compartments, measuring  $30 \times 30 \times 40$  cm (Fig. 1a). Thereby, two pairs of facing actor-partner compartments were created (left and right sides). The outer walls of the maze and the doors leading to the choice compartments

were opaque whereas the choice compartments themselves were separated from each other and from the opposite half of the maze by translucent walls containing an aluminum grid (approximately 80% open) in the lower half to facilitate sound transmission from the partner to the actor side, or vice versa. Instead of a social partner, in this jukebox experiment, ultrasonic speakers (Ultrasonic Dynamic Speaker Vifa, Avisoft Bioacoustics, Germany) were placed in each partner compartment to deliver acoustic stimuli at the vertical level of the actor animal's head at a distance of about 10 cm from the grid wall. As in the PCT, food rewards consisting of three sucrose pellets (45 mg dustless precision pellets, Bio-Serv, Germany) were delivered through a funnel into the choice compartment after playback of the acoustic stimuli.

### 2.3 Acoustic Stimuli

Three different types of acoustic stimuli were presented: 50-kHz USV stimuli, 22-kHz USV stimuli and background noise corresponding to the respective USV stimuli. All stimuli were presented with a sampling rate of 192 kHz in a 16-bit format for 5 s.

To determine whether the rat strain used to generate USVs mattered, or generalized across strains, we used two different sources of USV stimuli: type-1 stimuli were USVs recorded from Wistar rats, and described in detail by Wöhr and Schwarting (2007) and Sadananda, Wöhr & Schwarting (2008). Type-2 stimuli were based on calls recorded from pairs of interacting male Long-Evans rats. In brief, type-1 50-kHz USVs were recorded from a male Wistar rat exploring a cage containing scent from a cage mate. The stimulus consisted of 19 calls (total calling time: 1.19 s). Fourteen of these calls were frequency-modulated and five were flat. Call duration was  $0.06 \pm 0.01$  s (mean  $\pm$  SEM); peak frequency:  $61.41 \pm 1.51$  kHz; bandwidth:  $5.06 \pm 1.09$  kHz. The type-2 50-kHz calls were recorded during investigation of an unfamiliar juvenile conspecific by an adolescent rat. The stimulus consisted of 15 calls (total calling time: 1.47 s). Eleven of these calls were frequency-modulated and four were flat. Call duration was  $0.10 \pm 0.02$  s (mean  $\pm$  SEM); peak frequency:  $51.63 \pm 1.14$  kHz; bandwidth:  $6.09 \pm 1.35$  kHz. Eighteen different 50-kHz USV stimuli were generated by randomizing the order of the individual calls using SASLab Pro (version 5.2.08, Avisoft Bioacoustics, Glienicke, Germany). Background noise stimuli corresponding to the 50-kHz USV stimuli were generated by applying a band-rejection filter to eliminate the calls in the USV stimuli, leaving only background noise. The filter was set as to remove all signal components between 20.90 and 80.00 kHz. 50-kHz USV stimuli were played at approximately 69 dB and corresponding background noise was played at approximately 42 dB (measured from a distance of about 10 cm). Type-1 22 kHz calls were recorded from a male Wistar rat after applications of foot-shocks. Call duration was  $1.18 \pm 0.06$  s; peak frequency:  $23.61 \pm 0.07$  kHz; bandwidth:  $1.37 \pm 0.05$  kHz; type-2 22-kHz stimuli consisted of calls from another male adolescent Long-Evans rat investigating an unfamiliar juvenile conspecific. Eighteen USV stimuli with a duration of 5 s were generated by randomizing the order of 4 calls. The average duration of the calls was  $0.80 \pm$

0.07 s (mean  $\pm$  SEM) with a peak frequency of  $26.30 \pm 0.02$  kHz. Creation of corresponding background noise was similar to the 50-kHz stimuli, only now all signal components between 21.40 and 68.30 kHz and between 69.80 and 100.00 kHz were removed. The playback loudness was adjusted so that the ultrasonic components in the 22-kHz USV stimuli were played at approximately 69 dB. As such, the loudness of the background noise component was at approximately 32 dB.

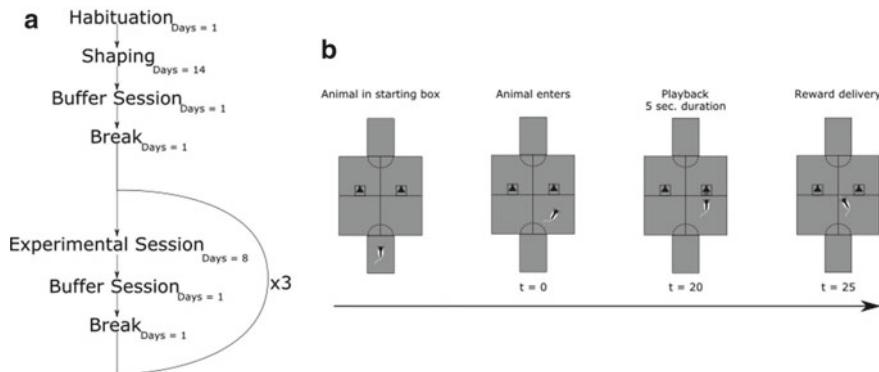
## 2.4 Task Design

Behavioral tests were performed under red light during the active period of the rats on consecutive weekdays. Before the beginning of the experiment, all rats received one day of habituation to the maze and 14 days of shaping sessions where they were gradually introduced to the testing conditions. Shaping procedures were similar to PCT training and consisted of daily sessions where animals acquired the trial structure (doors opening, compartment choice, doors closing, pellet delivery and consumption) up until the point where behavioral training was similar to the final test procedure except that no acoustic stimuli were presented.

In the final task, rats chose to enter one of the two choice compartments. Entering resulted in acoustic playback for five seconds. This 5 s USV playback period corresponds to the trial stage in the PCT when the partner is directed to the compartment facing the choice compartment with the actor animal, when the animals can interact acoustically through the aluminum grid. Ultimately, a food reward (three food pellets) was delivered to the actor rat, independent of which compartment was entered.

All rats performed the task under three conditions (Fig. 3), each for 8 consecutive sessions. Under condition 1 (50-vs-noise), a 50-kHz USV stimulus was played back in the choice compartment on one side and corresponding background noise in the choice compartment on the other side. Condition 2 (22-vs-noise) was identical to condition 1, except that a 22-kHz USV stimulus was presented together with corresponding background noise. In condition 3 (50-vs-22), the 50-kHz stimulus was played in one choice compartment and the 22-kHz stimulus was played in the other choice compartment. The order of experimental conditions was pseudo-randomized across rats within the groups, and the USV-compartment assignment was pseudo-randomized across days, ensuring that a given USV stimulus was not assigned to one side for longer than two days in a row. This pseudo-randomization approach was employed to mimic the pseudo-random assignment of the Both Reward option over days in the PCT, and to disambiguate playback preferences from potential side biases and habit development.

Each condition encompassed 8 daily testing sessions, which in turn consisted of four *forced trials* and 16 *free trials*. In the forced trials, only one door was opened in a pseudo-randomized order to allow rats to sample and learn the current assignment of acoustic stimuli to the choice compartments. In the free trials, both doors were opened at the same time and rats were able to choose which side to enter. Data is only reported for the free trials.



**Fig. 3** Sequence of the training and testing procedure and an individual experimental trial. **a** All animals went through habituation to the maze (Days = 1) and shaping (Days = 14), where they were gradually familiarized with the testing conditions. Afterwards, a buffer session (Days = 1; identical to the last shaping condition) took place, followed by a one-day break (Days = 1). Subsequently, rats were trained and tested in the final task (Days = 8), again followed by a buffer session (Days = 1) and a break (Days = 1). The procedure for the experimental sessions was repeated for all three conditions (curved arrow). **b** Before the beginning of a new trial, the animal was placed in the start box. Either one door (forced trials) or both doors (free trials) were opened, and, once the animal entered one of the two compartments, doors were closed and the trial timer was started ( $t = 0$  s). After a delay of twenty seconds ( $t = 20$  s), the USV stimulus was played back for five seconds in the respective compartment. Twenty-five seconds after trial onset ( $t = 25$  s), the food reward was delivered. After reward consumption rats were put back into their starting boxes for the next trial.

Figure 3b shows the sequence of an individual trial. In each trial, the animal is placed in the start box and the two doors leading to the choice compartments are opened. Once the animal enters one compartment, the doors are closed and the trial starts. After a delay of 20 s, the acoustic stimulus is played for 5 s. Subsequently, the food reward is delivered. After reward consumption, the animal is placed back into the start box for the next trial to begin. Adherence to the time points during each trial was ensured by a custom-made software script (Matlab 2014b, MathWorks Inc., USA) that also initiated the playback of acoustic stimuli (Avisoft-recorder, Avisoft Bioacoustics, Germany). After a session was finished, the maze was cleaned with a 70% ethanol solution to remove dirt and odor cues.

Both groups of rats, group 1 (USV<sub>Type-1</sub>) and group 2 (USV<sub>Type-2</sub>) performed this task; as described above, the only difference between the groups was the origin of the acoustic stimuli.

## 2.5 Data Analysis

Anticipating a transient response to the USV stimuli (Wöhr & Schwarting, 2012), we took advantage of the expected decay in preference both within and across sessions by

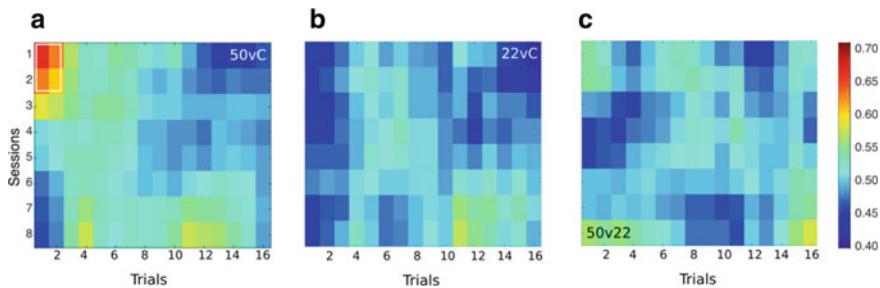
using a cluster-based permutation test derived from EEG/MEG/LFP time-frequency and spatio-temporal analysis included in the FieldTrip analysis Toolbox (Oostenveld et al. 2011). Briefly, in cluster permutation analysis, voxels (in our case, units of session-trial such as for example S3-T4) are assessed for significance by comparing the playback preference (choices of USV) across rats for that session-trial combination to a randomly permuted ( $N = 1000$  times) choice matrix (shuffling the position of USV choices but not the proportion). A reference distribution of preferences scores was constructed by averaging across rats for each session-trial unit across the preference scores resulting from the randomly permuted datasets and collection of these averages. Units of session-trial in the original dataset were flagged as significant if they fell outside the 99% confidence interval of this reference distribution. Clustering then took place by including adjacent significant units in a larger cluster (criterion: next-door-neighbours in horizontal (trial) or vertical (session) dimensions). The cluster statistic that resolves the multiple-comparison problem is computed by comparing the summed preferences for this cluster with the highest preference-sum of any cluster generated per random iteration (i.e. 1000 max-sum clusters). If, for positive (negative) clusters, the summed cluster score is higher (lower) than the 2.5% tail of the random cluster scores, the cluster as a whole is flagged as significant.

Following the analysis convention established for the PCT by Hernandez-Lallement, van Wingerden, et al. (2017), we also subdivided each session into three blocks of five trials and computed the mean compartment preference across trials within each block to contrast preferences between blocks. Analyses were performed using Matlab (2014b, MathWorks Inc., USA).

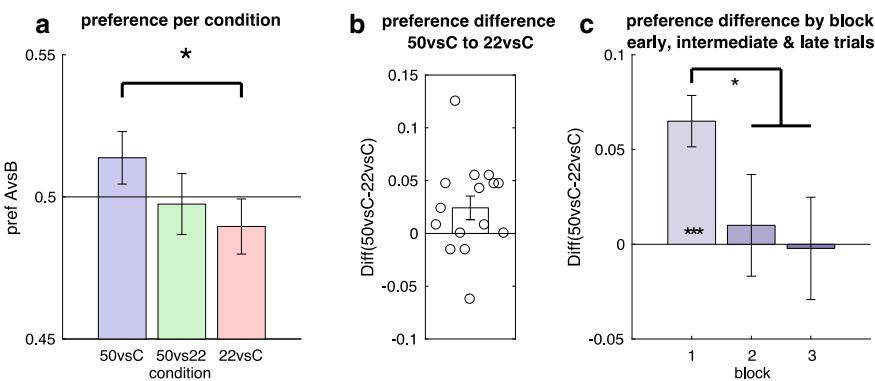
### 3 Results

As expected from the USV playback literature, we found a transient preference for the 50 kHz playback in the 50-vs-noise condition (Fig. 4a) and a transient preference against the 22 kHz playback in the 22-vs-noise condition (Fig. 4b). Cluster-permutation analysis indicated a significant  $2 \times 2$  cluster spanning sessions 1–2  $\times$  trials 1–2 ( $p < 0.05$  cluster permutation test, outlined in a white rectangle) in favor of the 50 kHzUSVs in the 50-vs-noise condition, while the transient preference against the 22-kHz playback in the 22-vs-noise condition visible in early trials across sessions did not reach statistical significance. Surprisingly, the sessions offering a direct choice between 50- and 22-kHz USV stimuli did not replicate this pattern. Instead of exhibiting a clear preference, rats were mostly indifferent between the 50- and 22-kHz USV playback (Fig. 4c), suggesting the possibility of an interaction of the call types when presented in the same setting.

This observation was supported by a more standard analysis, confirming that rats chose the compartment associated with USV stimulation significantly more often in the 50-vs-noise than the 22-vs-noise condition with both stimulus classes (paired-sample t-test,  $t_{(14)} = 2.16$ ,  $p < 0.05$ ; Fig. 5a). We observed inter-individual differences between the preference strengths for 50-kHz USVs (50-kHz vs. control) and 22-kHz



**Fig. 4** Preference maps for the three conditions, calculated for each session-trial unit, averaged across rats and smoothed using a 3-unit kernel. PseudoColor scale indicates level of preference for stimulus A (hot colors) vs stimulus B (cool colors). **a** 50 kHzUSV versus control, **b** 22 kHz USV versus control, **c** 50 kHz versus 22 kHz USV. White rectangle: significant preference cluster ( $p < 0.05$  cluster permutation test, corrected for multiple comparisons)



**Fig. 5** Preference difference for 50- over 22-kHz USV s when paired with its control stimulus, **a** preference per condition considering both stimuli types, all sessions and all trials, **b** difference in preference considering both stimuli types, all sessions and all trials. Barplots indicate mean difference in preference for the 50-kHz versus Noise minus preference for 22-kHz versus Noise,  $\pm$ SEM. Dots represent individual rats. **c** Same as in **b**, but now broken up in three blocks of five trials (trials 1–5, 6–10 and 11–16)

USVs (22-kHz vs. control; Fig. 5b). Blockwise-analysis, grouping trials 1–5, 6–10 and 11–16, showed that, in line with previous reports (Seffer, Schwarting, & Wöhr, 2014; Willuhn et al., 2014; Wöhr & Schwarting, 2007, 2012), the difference between the playback conditions was especially pronounced in the first block of five trials ( $6.5 \pm 1.4\%$ , tr. 1–5, one-sample t-test vs. 0;  $t_{(14)} = 4.80$ ;  $p < 0.001$ , Fig. 5c), as compared to blocks 2 ( $1.0 \pm 2.7\%$ , tr. 6–10;  $t_{(14)} = 0.37$ ; n.s.) and 3 ( $-0.2 \pm 2.2\%$ , tr. 11–16;  $t_{(14)} = -0.10$ ; n.s.). Indeed, the difference in preference in block 1 was significantly larger than the preference differences of blocks 2–3 combined (paired-sample t-test;  $t_{(14)} = 2.88$ ;  $p = 0.01$ ), confirming the transient nature of the effectiveness of USV playback in influencing spatial preferences.

Such a pattern of results could stem from either a preference for 50-kHz USVs over control stimuli, an avoidance of 22-kHz USVs over control stimuli, or both. Comparing the preference in the first block to the rest of the session suggests that only the preference for 50-kHz USVs over control was significantly higher in the first block (53.8 vs. 50.5%;  $t_{(14)} = 2.41$ ;  $p < 0.05$ ) while no differences could be detected in the 22-kHz USVs vs control condition (47.3 vs. 49.8%;  $t_{(14)} = -1.14$ ; n.s.).

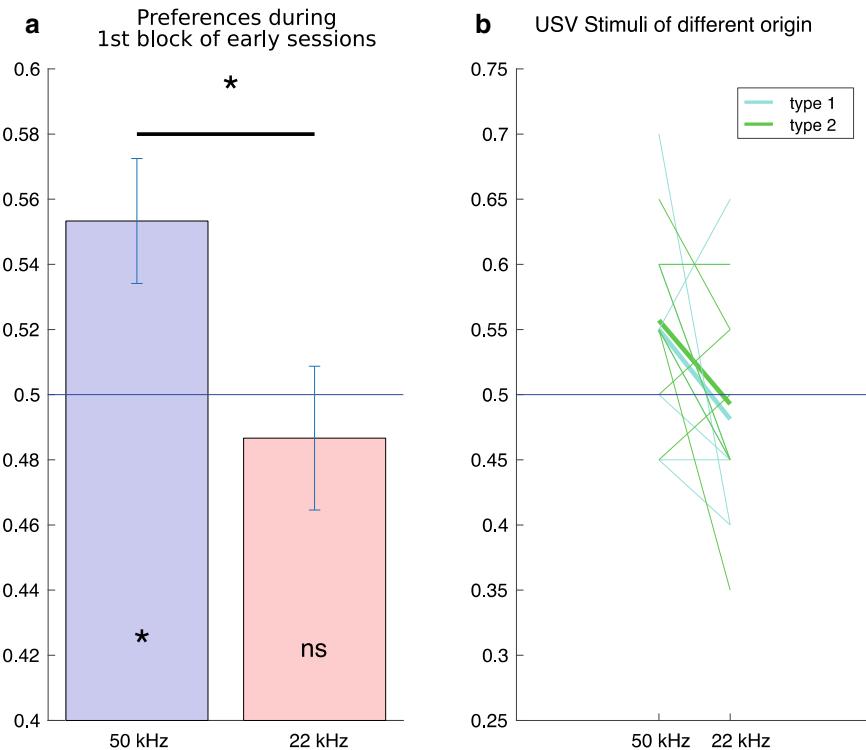
To gain further insights into the temporal pattern of the preference habituation effects and directly compare the effects found through the cluster based permutation approach, we compared preference for compartments in first trial block of the first half of sessions (sessions 1–4) with preferences in the second half of sessions (sessions 5–8). Interestingly, though some attenuation in preference across sessions could be found, the preference difference between the 50-vs-Noise and 22-vs-Noise condition for the first block showed up in the first half ( $6.7 \pm 2.8\%$ ,  $t_{(14)} = 2.43$ ;  $p < 0.05$ ) and the second half ( $6.3 \pm 2.7\%$ ,  $t_{(14)} = 2.35$ ;  $p < 0.05$ ) of sessions. However, only in the first half of the sessions did the preference in the first block of trials differ significantly from indifference in the 50 versus control condition ( $55.3 \pm 1.9\%$ ,  $t_{(14)} = 2.78$ ,  $p = 0.01$ , Fig. 6a).

Taken together, these results confirm that rats exhibit a transient preference for playback of 50-kHz USVs over non-ultrasonic control stimuli, combined with a trend towards avoidance of 22-kHz USV playback. As such, it seems plausible that USVs could be one channel of social feedback involved in driving spatial preferences linked to social outcomes in the PCT.

Finally, we asked if our Long-Evans rats responded differently to USVs originating from Long-Evans conspecifics (USV type-2 calls), or from rats from a different strain (Wistar rats; USV type-1 calls). However, our results showed that the pattern of results did not significantly differ between the USV-types used (Fig. 6b, independent samples t-tests at the level of 50 kHz playback, 22 kHz playback or the difference; all  $|t_{(13)}| < 0.25$ ; all  $p > 0.05$ ), suggesting that there is no evidence that rats discriminate between the strains of the USV sources.

## 4 Discussion

In this article, we present evidence supporting our hypothesis that USVs could act as social reinforcers, driving spatial preferences as observed in the pro-social choice task. In line with the social reinforcement hypothesis (Hernandez-Lallement, van Wingerden, et al., 2017), we theorized that USVs reinforce behavior that is associated with USV playback, but acoustic stimuli in a similar frequency range, yet without the social significance of USVs, do not act as social reinforcers. More specifically, we expected that 50-kHz USVs act as positive reinforcers, and that the probability of repeating actions coupled to 50-kHz USVs playback is larger than the probability of repeating actions associated with 22-kHz USVs or a non-ultrasonic control stimulus (Burgdorf et al., 2008). By contrast, we predicted that 22-kHz USVs act as negative



**Fig. 6** Preferences in 50-vs-Noise and 22-vs-Noise sessions, averaged across rats for the first block of five trials (1–5) and the first half of sessions (1–4). **a** Preference for 50 over noise was significantly above chance, while no significant difference from chance could be detected in the 22-vs-Noise sessions. The difference in preference for both session types was significant, though. **b** Individual data points for the data in **a**, now also split by stimulus type. No difference between stimulus type 1 (blue) and stimulus type 2 (green) could be detected

reinforcers, and that the probability of repeating actions associated with 22-kHz USVs is lower relative to 50-kHz USVs or non-ultrasonic control stimuli. Using an experimental paradigm adapted from the rodent PCT, we confirmed the reinforcing quality of USV playback, most prominent in the preference exhibited by rats for the playback of appetitive 50 kHz USV calls over control acoustic stimuli. The reinforcing quality is transient, however, as predicted from the literature (see below). Finally, we used two different sets of stimuli to test our hypothesis: one set of USVs was recorded from Wistar rats (Wöhr & Schwarting, 2007) and the other set from Long-Evans rats, as described above. We found that Long-Evans rats did not respond differently to USVs originating from conspecific Long-Evans rats, or from a different strain—Wistar rats.

Previous studies showed that 50-kHz USV stimuli induce strong, but transient approach behavior during initial playback and that this approach response quickly attenuated across trials (Wöhr & Schwarting, 2012; Seffer et al., 2014), together with

a decline in physiological measures of the rewarding properties of the USV stimulus (Willuhn et al., 2014). The authors explained this effect by USVs being secondary reinforcers that, after repeated exposure might, at least partially, lose their value (Willuhn et al., 2014). This explanation is in line with our hypothesis that USVs are not rewarding or aversive by themselves, but only by their virtue of carrying social significance in a social context.

A further issue that warrants elaboration is the nature of the motivating property of the USV stimulation. Because the USV playback stimuli were consistently paired with food rewards, as was the case in the partner session in the rodent PCT, we cannot conclude with certainty that USV playback by itself motived approach or avoidance behavior in the present study. Rather, the USV stimuli might have modulated the reinforcing value of the food rewards; that is the appetitive value of the food rewards was possibly enhanced by pairing it with 50-kHz playback and it was possibly reduced by pairing it with 22-kHz playback. Such a putatively modulating, rather than activating, effect of the USV stimuli on motivation might explain the relatively mild and transient size of the effects reported here.

Finally, our Long-Evans rats showed identical behavior towards USV stimuli recorded from Wistar and conspecific Long-Evans rats. Taken together, these data support our hypothesis that 50-kHz USVs, in contrast to comparable, but non-social acoustic stimuli, act as positive social reinforcers that influence behavior and might, therefore, contribute to orchestrating social interaction between rats. Our findings corroborate and extend the results of a recent study that showed that rats show instrumental responses to produce 50-kHz USV playback in a non-spatial operant conditioning setup (Burgdorf et al., 2008). However, the evidence for a putative role of 22-kHz USVs as negative social reinforcers is less conclusive. This result suggests that positive, rather than negative social feedback might drive the spatial preferences linked to different social outcomes (partner also rewarded or not) in the pro-social choice task.

Although the social reinforcement mechanism described here and elsewhere (Hernandez-Lallement, van Wingerden, et al., 2017) provides a parsimonious, plausible and realistic explanation for rat social behavior, it is agnostic about how rats actually cognitively represent their social world: as discussed above, our social reinforcement theory does not explain how a rat conceptually links and represents the different stimulus levels—the USV's physical dimension (rhythmic oscillations of air compression and deflation creating auditory perception), the emotional level (the putative enjoyment or averseness of listening to USVs) and their motivational level (50 kHzUSVs are wanted and they prompt action to obtain them) – into a coherent cognitive representation of a social situation. In the following section, we will present a philosophically inspired attempt to theoretically model how rats link and process these stimulus levels into a complex cognitive representation of social interaction.

The second part of this paper, thus, attempts to provide a novel approach to animal learning and cognition. The “cascade” approach regards the categorization and cognitive representation of types of action as potentially multilevel. When a rat learns in an experimental setting that certain types of action are rewarding, its brain is assumed to form an action cascade that categorizes this type of action simultaneously

as an act of getting a reward. The multilevel approach can be applied to model social behavior as multilevel: a cognitive complex of performing the basic physical behavior and thereby at the same time a particular kind of social behavior. Applying likewise to human cognition, cascade theory is a candidate for connecting animal and human cognition.

## Part II: Cascades in Animal Cognition

### 5 A Cognitive Perspective: Acting at Multiple Levels

This section offers a theoretical perspective on the neurocognition of the representation of action. Applied to rats, it is not to be taken as a theory rival to existing psychological accounts of animal learning, but rather as an account concerning the cognitive representation involved and the cognitive implementation of conditioning. The most prominent feature of the “Cascade” theory of cognitive representation is a multilevel approach to categorization. It applies, it appears, to humans and animals likewise.<sup>1</sup>

#### 5.1 *Goldman’s Multilevel Theory of Human Action*

##### 5.1.1 Goldman’s Notion of Level-Generation and the Notion of Cascade

When humans categorize and conceptualize an action, they usually do it in more than one way at the same time. The philosopher Alvin Goldman developed a theory of human action that is based on this principal observation (Goldman, 1970). If I open a door, this is a physical act of interaction with an object that changes its state. Opening a particular door can be achieved by a variety of bodily actions. If it is a hinged door, I can push the door at its handle or somewhere else with my hand, I can push it with my foot, I can lean against it with my shoulder or my back; depending on my position and the construction of the door, I may have to pull at the door. For sliding doors or automatic doors, other types of action are required. Thus, ‘opening a door’ refers to at least two levels of action: (1) the basic physical action one applies to the door, and (2) the more abstract functional level of causing the door to open. The acts at the physical and at the functional level do not concern the same properties of the door. The physical act changes the spatial position of the door leaf or leaves. The higher-level act concerns states of the door that are related to its functioning as an object that is used to obstruct or enable access to a space behind it.

The lower-level action is necessary for achieving the higher-level action. This achievement is not automatic but requires certain circumstances; for example, the

---

<sup>1</sup>The theory is introduced in more depth and detail in Löbner ([this volume](#)).

mechanical door must not be locked, the automatic door must be in function. Goldman (1970) speaks of “level-generation” if actions are related in this way: under certain circumstances, the lower-level action “generates” the higher-level action, the lower-level action is a *method* of doing the higher-level thing; by pushing the door or pulling at it, one opens it. While in this case the level-generating relation is based on causation, there are also other mechanisms such as conventional level-generation; for example, if I nod my head, this may conventionally generate an approval or permission because nodding one’s head is a conventionally established method of approving or permitting.

Crucially, if an action A generates a higher-level action B, A and B are actions by the same agent and at the same time, done in one. It is very important to note that level-generation does not relate an action to an event it causes. If I open the door for someone and let them pass, I first open the door and then the other will pass through the door a moment later. Level-generation does not obtain between these consecutive actions by two different agents. Rather it obtains between the action of opening the door and the action of opening a passage *for the other*. These two actions are actions by the same agent and they occur at strictly the same time. It is this feature of Goldman’s theory of action that makes it a theory of multilevel categorization.

According to Goldman, a basic action may level-generate more than just one higher-level action; it can generate a complex multilevel structure of actions with many steps that build on each other; the structure can also branch into different lines of generation. For example, by pushing a door and opening it, one may at the same time open a passage in an aisle as well as cause an air draft; opening the passage may in turn generate doing a favor; causing a draft may further generate making a window slam. We will give complex examples below. Goldman uses the term “act-trees” for structures created by level-generation; we prefer to call them “cascades” as there are good reasons to transfer the notion to other things than action<sup>2</sup>. Crucially, the actions that form a cascade are actions of different type. For example, leaning against the door and opening the door are not actions of the same type. A door can be opened by other methods, and leaning against the door can have other effects than opening it; for example, it may as well be an act of closing the door, or of keeping the door closed if somebody is pushing against it from the other side.

### **5.1.2 Goldman’s Level-Theory as a Psychological Theory of Categorization**

It is convenient to use the term ‘doing’ for that to which a cascade description applies: there is *one doing*, for example with the door, but this one doing can be categorized in many different ways as constituting as many different types, or categories, of action as the cascade provides. In the discussion of his theory of action with other philosophers, Goldman emphasizes that the distinction of types involved in a cascade of action is a psychological distinction, not a distinction of things out there in the

---

<sup>2</sup>See Löbner, [this volume](#), ss. 5–7.

world. The cascade agent produces one doing, but it is categorized simultaneously at different levels in a hierarchy of level-generation (Goldman, 1979). A cascade forms in our minds, in our view and cognitive modeling of what is going on or what we are doing ourselves. What a person does in a concrete situation, to us, *is*, in our reality, all these acts in the cascade at the same time. The particular doing in our door example of level-generation may belong at the same time to the action categories ‘push against the door’, ‘open the door for Adam’, ‘do Adam a favor’, and maybe others. It is important to realize that the different categories we may apply to the one underlying doing are not just a bunch of categories that are somehow associated. Rather they are organized in a tree structure of dependence. The higher actions depend for their coming about on the lower actions that “generate” them. And all higher-level actions depend on necessary circumstances to come about.

As the door example illustrates, the formation of cascades takes place even with as simple actions as opening a door. We may well assume that humans categorize almost any willful action by a human as a *cascade* of action rather than just as the basic physical doing. We will inevitably try to interpret the actions of others in terms of the intentions they pursue by doing what they do; if they act on an artefact in a normal way, for example on a door, we will assume that the action is related to the usual function of the object. Thus, categorizing an action as ‘opening the door’ would provide a *causal explanation* of the observable physical act.

### 5.1.3 Social Action and Interaction

One observation relevant in our context is the fact that social action necessarily constitutes higher-level action. Searle (1995) developed a theory of social reality that distinguishes between a physical level and a social level of action, persons, and objects. A certain movement with the head is an approval if and only if it *counts as* such; a human is the president of Canada if and only if they *count as* such, and a piece of paper is money if and only if it *counts as* money. The things that count as something in these examples are *physical* entities and what they count as are *social*, entities that is, entities in our social reality. Notably, in all these cases, the things considered are necessarily both at the same time: the physical entity and the social-reality entity. For the part of Searle’s theory concerning acts, the relation between things at the physical and at the social level is captured by Goldman’s more general notion of level-generation.

As a consequence of the principal higher-level character of social action, social behavior always ‘parasitizes’ on more basic physical behavior.<sup>3</sup> For example, one may turn up the corners of one’s mouth and expose the front teeth and thereby level-generate a smile which, if directed at someone, may under circumstances constitute a social signal which constitutes a display of affection, or something else. Up from the level ‘smiling *at someone*’, the cascade reaches a social level. If we go back to

---

<sup>3</sup>The terminus *parasite* was introduced in this connection by Kearns (2003) who relates to the lower and the higher level of a two-level cascade as ‘host’ and ‘parasite’, respectively.

**Fig. 7** A door-opening cascade: doing six things in one

```

graph TD
    A1[A gains (a little) pleasure] -- "↑" --> A2[A makes B smile at A]
    A2 -- "↑" --> A3[A obliges B]
    A3 -- "↑" --> A4[A does B a favor]
    A4 -- "↑" --> A5[A keeps the door open for B]
    A5 -- "↑" --> A6[A keeps the door open]
  
```

A gains (a little) pleasure  
↑  
A makes B smile at A  
↑  
A obliges B  
↑  
A does B a favor  
↑  
A keeps the door open for B  
↑  
A keeps the door open

the example in the introduction, we get an even more complex structure. Using an upward arrow ↑ for level-generation, we can represent the cascade bottom-up as in Fig. 7.

As the example illustrates, own action may cascade to ultimately giving oneself a pleasure (or any other kind of emotional experience) by doing what one does. Obviously, this too cannot be done without the support of some physical action. We may keep in mind two general points about cascades: (i) physical action may cascade to social action, and (ii) action may cascade to obtaining an emotional experience, where emotional-level action may or may not come about by means of social-level action like in our fictitious example.

There is another aspect to the door-opening example. Social reality is constructed interactively (see Clark (1996) on a multilevel interactional account of verbal communication). If A keeps a door open for B, meaning to do B a favor and cascading the conceptualization of their own act correspondingly, then the thanks A receives from B will confirm that A and B share the social construal of their interaction: B would not have thanked A if B had not construed A's act as involving the level-generation of doing B a favor. An analogous consideration applies to the next level above the favor: the level-generation of obliging B by doing B a favor. Acknowledgment and confirmation of this additional, emotional level, is executed by B sending a smile to A. Given that receiving a smile is felt as something pleasant, the level-generation of 'B please A' by 'B smile at A' is part of the joint construal of B's reaction.

## 5.2 Cascades and Learning

Goldman's theory was constructed for the categorization of individual action tokens. It can, however, also be applied to the consistent multilevel categorization of recurring *types* of action. For example, if we experience that the light goes on when we flip a certain switch, and if we repeat the action and achieve the same effect, we will learn a cascade: that flipping this switch goes with switching that light on. We acquire a piece of procedural knowledge by memorizing a two-level action cascade concept composed of the two single-level action concepts 'flip this switch' and 'turn on that light'. Our environment being as it is, we will easily generalize this cascade to other switches and other lights, and so on. Thus, action learning is cascade learning, at least for all but merely physical basic action like turning one's head or lifting ones hand. We learn that doing one thing also means doing the higher-level thing, and the level above that, and so on. An action and the higher level achieved with it are conflated into one concept. Cascade learning may also include that, by an action, we trigger approval or disapproval, cause pleasure or pain, a particular taste or other bodily sensations. If we assume that cascading plays a role in concept formation, we may conclude that action concepts are formed that link basic actions and the recurring achievement of certain causal effects into one multilevel concept.

It is important to note that even for humans, kids or adults, learning of action cascades does not necessarily involve reflection. It just requires that the learning subject register that the lower-level action goes with the higher-level action. In particular, the learning of cascade levels that are causally linked does not require any *causal understanding*. We learn that pressing the red button of the TV remote control means turning the TV on or off, but we may well die without ever having understood what we actually do at the technical level by pressing this button. This level of understanding is not relevant for learning how to succeed in dealing with TVs and remote controls. To know how to deal with a remote control is essentially 'knowledge how'<sup>4</sup>, and the mechanism by which we acquire this knowledge is learning by doing.

Cascade learning does not only concern practical abilities. A child may cascade-learn that a certain kind of behavior always upsets her mother; the child will register this and adjust her behavior accordingly, but may possibly never understand *why* her mother reacted that way. We learn in countless regards that our actions are accompanied by higher cascade levels of particular qualities. Cascade learning will result in a "practical" implicit understanding of the environment, in the sense that we learn which intended or unintended higher-level kinds of action are generated by certain other kinds of action. We learn things like "if I do x, I give myself experience y". Given that we are able to undertake certain action or refrain from it, this kind of understanding our environment will enable us to adapt to it.

Cascade knowledge need not be accessible to consciousness: we may have it without being aware of it and without being able to describe it. For example, pronouncing a word in a way that enables others to recognize it phonetically means to enact a cascade of production based on intentional action of our articulatory organs

---

<sup>4</sup>See Katzoff (1984) for the connection of knowing how to Goldman's theory of action.

to produce articulated sounds, thereby producing speech sounds, thereby producing certain speech phonemes, and with them an established sound form of a word in a particular language. All this is stored in the language production repertoire—a normal language user is not aware of the levels of actions involved and they would not be able to describe what they do at which levels. All they need to be able to do is to aim at doing something particular at a pretty high level, something with the result of making audible a particular sound pattern.

### ***5.3 Applying Cascade Theory to Rat Behavior in the Experiments Reported***

We proceed to propose that the cascade model of action categorization and action learning applies to rats as well. First of all, it appears that there are certain types of rat action that are relevant to the actors at levels beyond the mere physical doing. Among these are levels that constitute social action. For example, if young rats do rough-and-tumble play, they recognize that this is not hostile fight: crucially, the fight is ‘friendly’ *to both of them*. In some way or other, they succeed in letting the other “know” that their own behavior is not hostile, and they succeed in categorizing the other’s behavior in the same way. Both rats engaged in a rough-and-tumble play possess two categorizations of representing physical fight or else fight-like action. At a lower level they categorize the physical action, at a higher, social, level they categorize it as hostile fighting or as play. At the lower level they “know” bodily methods of fighting, for instance, pushing or biting, and they are able to modulate these methods as to cascade either to a real fight or to rough-and-tumble play. There can be no doubt that a rough-and-tumble play *to the rats* is both, bodily interaction and a social interaction that is different from hostile fight. What they do has a function to them at both levels, as some sort of bodily learning and some sort of social learning.

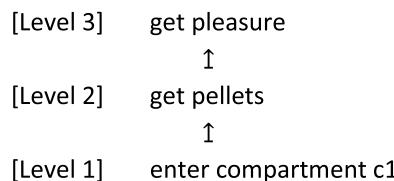
When we say that this “is to the rats” a particular type of action, we do not imply consciousness on the part of the rat. The cascade view does not commit us to the assumption that rats have consciousness (at least not in the same sense as humans); it only commits us to assume that the rats’ brains categorize the rats’ doing in these ways and that, in this sense, the rats register what is going on at both levels. As rats are able to recognize and repeat *types* of action, for example under experimental conditions, they must have cognitive representations of types of action. Crucially, they register the character of what is going on not only for their own part, but also for the part of their interaction partner.

Among the actions that have multilevel character to rats are the USVs (ultrasonic vocalizations) mentioned above. The fact that these vocalizations trigger brain reactions associated with emotion, shows that these are not just plain sound productions (like, for example, the production of the sound they produce when they scratch their ear or shuffle around); these special sound productions are ‘received’ at an

acoustic *and* an emotional level. We do not know if the rat, when hearing a 50-kHz USV, hears this as a *display* of comfort or pleasure. If this should be the case, the rat might have a two-level representation of the act *by their conspecific*. All we seem to be entitled to assume at present is that 50-kHz USVs must have a pleasant ‘ring’ to the perceiver. But this is sufficient for the assumption of a two-level neural cascade representation of 50-kHz USVs issued by other rats, whence these USV’s carry emotional significance.

In the experiments, the rats learn. They acquire behavior. The experiments are designed in the way that the behavior acquired leads to getting themselves a reward. We can apply the cascade model to the learning process, if we assume that learning a particular behavior consists in acquiring a multilevel action cascade. The general structure of reward-inforced learning would be the acquisition of a cascade that amounts to: ‘do x’ ↑ ‘get a reward’; here ‘get’ is to be taken to mean active acquisition, not just passive reception, because the latter would not be an action by the animal.

Assume that the actor rat learns that it will receive pellets upon entering compartment c1. That learnt, the rat will repeat the action if it likes to get pellets. This behavior can be interpreted as involving the acquisition of an action cascade of three levels:

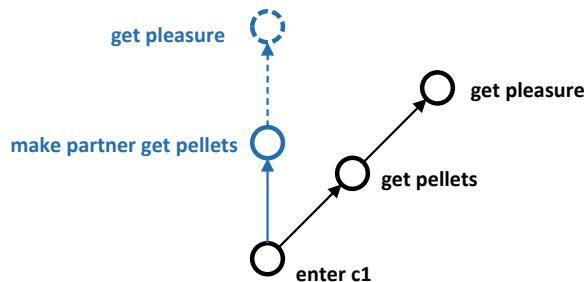


One might speculate that it is the rewarding course of events that supports not just the behavior as such, but primarily the formation of the cascade described; if the animal forms and then memorizes the cascade, this results in a mental condition that enables the animal to repeat these rewarding experiences at will by taking the action at the bottom of the cascade.

In the prosocial choice task experiments described in Hernandez-Lallement et al. (2015), some rats seem to have learnt just Cascade 1. The prosocial rats, however, developed a behavior that involves a more complex cascade structure with a second branch on the first node (Fig. 8, blue branch).

They register that the partner rat gets pellets, too, and their brain ascribes it to themselves as a generated higher-level action. As for the third step of the cascade, we know that there are 50 kHz USVs when the actor rat and the partner rat simultaneously get their pellets; however, due to the technical equipment used, it was not possible to ascribe the vocalizations to one or the other rat or to both. We are entitled to assume that the actor rat sees and thereby registers that the partner rat gets pellets. We do not know whether this constitutes a pleasant experience to the actor rat. If we could be sure that the partner rat produces a 50 kHz USV, we might assume that the actor rat hears it and experiences this as an emotional reward. We can explain the preference for this condition only if we assume that the prosocial behavior cascades

**Fig. 8** Act-trees for the actions taken by the rats in the PCT experiment. The black branch represents the action cascade acquired by every rat in this experiment. The blue structure represents the additional cascade branch acquired by the rats that develop prosocial behavior



to an additional reward in the left branch of the cascade. The left cascade branch would then level-generate an additional third-level ‘get pleasure’.

In the new experiments described above, with no partner rat present, the USV constitutes an additional two-step branch generated by Level 2 in the first cascade, to be construed as ‘get a 50 kHzUSV’, thereby ‘get pleasure’. We will assume that the two-way reward (transiently) outweighs the one-way reward of the no-USV condition. An explanation as to why the effect of the USV gets weaker in the course of the experiment will not be attempted here.

#### **5.4 Psychological Commitments of the Cascade Approach**

Application of the cascade model to rat learning involves certain psychological commitments.

(i) The rat's brain implements cascade formation.

The rat's brain creates links between basic types of bodily action and what goes with them, perceptibly to the rat; if the rat brain works in this way, it ascribes the effects of behavior to the behavior itself, connecting, for example, eating certain food to staying hunger. In this way, the animal learns by experience what its behavior "means" to it. When we talk of "meaning" here, we mean it in the basic sense of immediate concomitance, not involving reasoning or convention: if something is of category A and category A cascades to category B, then this instance of A "means"/"constitutes"/simply "is" also an instance of B—to the cognitive subject.

Level-generation presupposes that the animal perceives its own action, and attributes it to itself. This results in a second psychological commitment:

(ii) Rats have a (weak) sense of agency. Their brain records their action.

There can be no doubt that rats, by way of proprioception and perception of the environment, sense that they are acting.

In addition, we need to assume that the rat's brain *categorizes* what the animal is doing. This amounts to the following commitment:

(iii) The rat's brain forms concepts (representations) of types of own action.

Crucial for the cascade formation is the following assumption:

- (iv) The rat's brain assigns credit to the animal itself for what happens concomitant with action of the particular type.

Cascade formation then means that the rat's brain generates a higher cascade level for the underlying action concept that amounts to making happen what happens after action of the given type.

There are restrictions on this condition. First, we will assume that it holds only for such events following the rat's action that are significant to the animal's well-being, and hence of "interest". Second, there is supposed to be a limit on the time that may lapse between the rat's action and later events. The rat's brain will possibly not connect the animal's doing to things that happen after a long time.

It appears that commitments (ii) to (iv) are uncontroversial; we construe the changes in behavior of rats in experimental settings as learning behavior under the conditions of the experiment. This would be unexplainable if we would not assume that the animals' brains register the animals' doings as their own and as of a particular type and if their brains did not credit the animals with what follows their own action as something they can ascribe to this type of action.<sup>5</sup>

- (v) The rat's brain stores in long-term memory the repeated concomitance of certain effects with a type of own action.

This means that the rat's brain connects this *type* of action—not only individual single action tokens—with this *kind* or result.

Of course, the crucial assumption is the first one. The other assumptions are implicit in everyday experimental practice.

## 5.5 What Can the Cascade Approach Buy Us?

What the cascade theory buys us is twofold. First, it provides a fundamental neurocognitive mechanism for a model of the animal's learning about its environment. If the animal's brain builds cascades on the types of physical action the animal is capable of, then the brain integrates the type of action with the achievement of its results into one multilevel concept. The type of action is thereby invested with a particular significance for the animal, for example emotional significance, significance relevant for survival, or the significance of performing a certain type of social action or interaction. Cascade-format action concepts link an action to the achievement of its result as something ascribed to the animal as self-caused—and thereby controlled by own behavior. Cascade learning of effects of their doing invests the animals with the ability to choose ways of action, to seek advantage and avoid disadvantage. Thus, cascade formation for own-action types provides a basic mental mechanism of adapting to the environment, including the animal's social group, in a learning-by-doing way.

---

<sup>5</sup>See Takahashi et al. (2011) for exceptional conditions of the animals under which the credit assignment required does not work.

Second, the cascade approach offers an explanation for the way in which an animal is able to acquire a *practical* understanding of the ways of its environment, as its brain links types of behavior to the triggering of its outcome. If the cognitive system of an animal is equipped with the ability of action cascade formation—i.e. if it records what the animal does to itself if it acts in this way or another—it enables adaptation to the environment without requiring any level of causal understanding, reasoning, or modeling. Thus, the cascade model of learning is a model of learning by doing and what is acquired is plain knowledge-how.

The cascade approach might be successful in modeling multilevel categorization across humans and animals, in particular as part of modeling the acquisition of multilevel action concepts and methods of how to do things, and of what is ‘social reality’ to the cognitive subjects. Another way of looking at cascade theory is to consider it a psychological theory of “meaning”, in the very basic sense that acting at a lower cascade level also “means” to act at the generated higher level. In this sense, cascading provides action with meaning to the cognitive subject.

In the field of cognitive theory and psychology, the theory is at its very beginning. It seems to be able to claim some plausibility (cf. Vallacher & Wegner, 2011). In any event it would be interesting to try to develop methods for testing it experimentally. For example, a cascade approach to learning raises concrete questions concerning structural constraints on cascades to be acquired in terms of the number of levels, of branching complexity, and of memorizability.

## 6 Conclusions

In this study, we present evidence supporting our hypothesis that USVs act as social reinforcers. In line with the social reinforcement hypothesis (Hernandez-Lallement, van Wingerden, et al., 2017), we show that rats preferred T-maze compartments associated with 50-kHz USV playback over compartments associated with non-ultrasonic control stimuli. This observation fuels the hypothesis that USVs might orchestrate and structure social interaction between rats. Finally, we argue that one avenue towards understanding the conceptual representation of the emotional and motivational significance of rat USVs might require a multilevel approach, as proposed by Goldman (1970) in his cascade model of mental representation of human action.

**Acknowledgements** This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft), projects B09 and D03 of CRC 991 “The structure of representations in language, cognition, and science”. Mireille van Berkel is supported by Volkswagen Freigeist fellowship no. AZ88216. Maurice-Philipp Zech is supported by a grant from the German Research Foundation (grant no. KA 2675/5-3).

## References

- Ben-Ami Bartal, I., Decety, J., & Mason, P. (2011). Empathy and pro-social behavior in rats. *Science*, 334, 1427–1430.
- Blanchard, D. C., & Blanchard, R. J. (1990). Behavioral correlates of chronic dominance-subordination relationships of male rats in a seminatural situation. *Neuroscience and Biobehavioral Reviews*, 14, 455–462.
- Blanchard, R. J., Flannery, K. J., & Blanchard, D. C. (1988). Life-span studies of dominance and aggression in established colonies of laboratory rats. *Physiology & Behavior*, 43(1), 1–7.
- Brudzynski, S. M. (2013). Ethotransmission: communication of emotional states through ultrasonic vocalization in rats. *Current Opinion in Neurobiology*, 23(3), 310–317.
- Brudzynski, S. M., & Ociepa, D. (1992). Ultrasonic vocalization of laboratory rats in response to handling and touch. *Physiology & Behavior*, 52(4), 655–660.
- Burgdorf, J., Knutson, B., Panksepp, J., & Ikemoto, S. (2001). Nucleus accumbens amphetamine microinjections unconditionally elicit 50-kHz ultrasonic vocalizations in rats. *Behavioral Neuroscience*, 115(4), 940–944.
- Burgdorf, J., Kroes, R. A., Moskal, J. R., Pfau, J. G., Brudzynski, S. M., & Panksepp, J. (2008). Ultrasonic vocalizations of rats (*Rattus norvegicus*) during mating, play, and aggression: Behavioral concomitants, relationship to reward, and self-administration of playback. *Journal of Comparative Psychology*, 122(4), 357–367.
- Calvino, B., Besson, J. M., Boehrer, A., & Depaulis, A. (1996). Ultrasonic vocalization (22–28 kHz) in a model of chronic pain, the arthritic rat: effects of analgesic drugs. *NeuroReport*, 7(2), 581–584.
- Chang, S. W., Fagan, N. A., Toda, K., Utevsky, A. V., Pearson, J. M., & Platt, M. L. (2015). Neural mechanisms of social decision-making in the primate amygdala. *Proceedings of the National Academy of Sciences*, 112(52), 16012–16017.
- Clark, H. H. (1996). *Using Language*. Cambridge University Press.
- Engelhardt, K. A., Fuchs, E., Schwarting, R. K., & Wöhr, M. (2017). Effects of amphetamine on pro-social ultrasonic communication in juvenile rats: Implications for mania models. *European Neuropsychopharmacology*, 27(3), 261–273.
- Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960), 785–791.
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition and cooperation. *The Quarterly Journal of Economics*, 114, 817–868.
- Fiske, S. T., & Taylor, S. E. (1984). *Social Cognition*. Reading and MA: Addison-Wesley Publishing.
- Goldman, A. I. (1970). *Theory of Human Action*. Princeton, New Jersey: Princeton University Press.
- Goldman, A. I. (1979). Action, Causation, and Unity. *Nous*, 13, 261–270.
- Hernandez-Lallement, J., van Wingerden, M., Marx, C., Srejic, M., & Kalenscher, T. (2015). Rats prefer mutual rewards in a prosocial choice task. *Frontiers in Neuroscience*, 8, 443.
- Hernandez-Lallement, J., van Wingerden, M., Schäble, S., & Kalenscher, T. (2016). Basolateral amygdala lesions abolish mutual reward preferences in rats. *Neurobiology of Learning and Memory*, 127, 1–9.
- Hernandez-Lallement, J., van Wingerden, M., & Kalenscher, T. (2017). Towards an animal model of callousness. *Neuroscience and Biobehavioral Reviews*.
- Hernandez-Lallement, J., van Wingerden, M., Schäble, S., & Kalenscher, T. (2017b). A social reinforcement learning hypothesis of mutual reward preferences in rats. *Current Topics in Behavioral Neurosciences*, 30, 159–176.
- Ishiyama, S., & Brecht, M. (2016). Neural correlates of ticklishness in the rat somatosensory cortex. *Science*, 354(6313), 757–760.
- Janak, P. H., & Tye, K. M. (2015). From circuits to behaviour in the amygdala. *Nature*, 517(7534), 284–292.
- Kalenscher, T., & van Wingerden, M. (2011). Why we should use animals to study economic decision making: A perspective. *Frontiers in Neuroscience*, 5(82), 1–11.

- Kashtelyan, V., Lichtenberg, N. T., Chen, M. L., Cheer, J. F., & Roesch, M. R. (2014). Observation of reward delivery to a conspecific modulates dopamine release in ventral striatum. *Current Biology*, 24(21), 2564–2568.
- Katzoff, C. (1984). Knowing how. *The Southern Journal of Philosophy*, 22, 61–68.
- Kearns, K. (2003). Durative achievements and individual-level predicates on events. *Linguistics and Philosophy*, 26(5), 595–635.
- Knutson, B., Burgdorf, J., & Panksepp, J. (1998). Anticipation of play elicits high-frequency ultrasonic vocalizations in young rats. *Journal of Comparative Psychology*, 112(1), 65–73.
- LeDoux, J. E. (1994). Emotion, memory and the brain. *Scientific American*, 270(6), 50–57.
- LeDoux, J. E., Cicchetti, P., Xagoraris, A., & Romanski, L. M. (1990). The lateral amygdaloid nucleus: sensory interface of the amygdala in fear conditioning. *Journal of Neuroscience*, 10(4), 1062–1069.
- Löbner, S. (this volume). Cascades. Goldman's level-generation, multilevel categorization of action, and multilevel verb semantics.
- Lukas, M., & Wöhr, M. (2015). Endogenous vasopressin, innate anxiety, and the emission of pro-social 50-kHz ultrasonic vocalizations during social play behavior in juvenile rats. *Psychoneuroendocrinology*, 56, 35–44.
- Oberliessen, L., Hernandez-Lallemand, J., Schäble, S., van Wingerden, M., Steinstra, M., & Kalenscher, T. (2016). Inequity aversion in rats, *Rattus norvegicus*. *Animal Behaviour*, 115, 157–166.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, 156869, 9. <https://doi.org/10.1155/2011/156869>.
- Panksepp, J., & Burgdorf, J. (2000). 50-kHz chirping (laughter?) in response to conditioned and unconditioned tickle-induced reward in rats: Effects of social housing and genetic variables. *Behavioural Brain Research*, 115(1), 25–38.
- Parkinson, J. A., Robbins, T. W., & Everitt, B. J. (1996). *Lesions of the nucleus accumbens core, but not basolateral amygdala or subiculum, disrupt stimulus-reward learning in a novel autoshaping procedure*. Paper presented at the Society for Neuroscience Abstracts.
- Parsana, A. J., Li, N., & Brown, T. H. (2012). Positive and negative ultrasonic social signals elicit opposing firing patterns in rat amygdala. *Behavioral Brain Research*, 226(1), 77–86.
- Rutte, C., & Taborsky, M. (2007). Generalized reciprocity in rats. *PLoS Biology*, 5(7), e196.
- Sadananda, M., Wöhr, M., & Schwarting, R. K. (2008). Playback of 22-kHz and 50-kHz ultrasonic vocalizations induces differential c-fos expression in rat brain. *Neuroscience Letters*, 435(1), 17–23.
- Sales, G. D. (1972). Ultrasound and aggressive behaviour in rats and other small mammals. *Animal Behaviour*, 20(1), 88–100.
- Searle, J. R. (1995). *The construction of social reality*. New York: The Free Press.
- Seffer, D., Schwarting, R. K., & Wöhr, M. (2014). Pro-social ultrasonic communication in rats: Insights from playback studies. *Journal of Neuroscience Methods*, 234, 73–81.
- Siviy, S. M., & Panksepp, J. (2011). In search of the neurobiological substrates for social playfulness in mammalian brains. *Neuroscience and Biobehavioral Reviews*, 35(9), 1821–1830.
- Strombach, T., Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, I. I., Tobler, P. N., et al. (2015). Social discounting involves modulation of neural value signals by temporoparietal junction. *Proceedings of the National Academy of Sciences*, 112(5), 1619–1624.
- Sutton, R. S., & Barto, A. G. (2012). *Reinforcement Learning: An Introduction*. Cambridge, Massachusetts: The MIT Press.
- Takahashi, Y. K., Roesch, M. R., Wilson, R. C., Toreson, K., O'Donnell, P., Niv, Y., et al. (2011). Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nature Neuroscience*, 14(12), 1590–1597.
- Vallacher, R. R., & Wegner, D. M. (2011). Action identification theory. In P. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories in social psychology* (pp. 349–370) London, England: Sage.

- Vanderschuren, L. J., Achterberg, M. E., & Trezza, V. (2016). The neurobiology of social play and its rewarding value in rats. *Neuroscience and Biobehavioral Reviews*, 70, 86–105.
- Willuhn, I., Tose, A., Wanat, M. J., Hart, A. S., Hollon, N. G., Phillips, P. E., et al. (2014). Phasic dopamine release in the nucleus accumbens in response to pro-social 50 kHz ultrasonic vocalizations in rats. *Journal of Neuroscience*, 34(32), 10616–10623.
- Wöhr, M., & Schwarting, R. K. (2007). Ultrasonic communication in rats: can playback of 50-kHz calls induce approach behavior? *PLoS ONE*, 2(12), e1365.
- Wöhr, M., & Schwarting, R. K. (2012). Testing social acoustic memory in rats: effects of stimulus configuration and long-term memory on the induction of social approach behavior by appetitive 50-kHz ultrasonic vocalizations. *Neurobiology of Learning and Memory*, 98(2), 154–164.
- Wöhr, M., & Schwarting, R. K. (2013). Affective communication in rodents: ultrasonic vocalizations as a tool for research on emotion and motivation. *Cell and Tissue Research*, 354(1), 81–97.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Influence of Manner Adverbs on Action Verb Processing



Jan Sieksmeyer, Anne Klepp, Valentina Niccolai, Jacqueline Metzlaff,  
Alfons Schnitzler, and Katja Biermann-Ruben

**Abstract** Language-motor interaction is suggested by the involvement of motor areas in action-related language processing. In a double-dissociation paradigm we aimed to investigate motor cortical involvement in the processing of hand- and foot-related action verbs combined with manner adverbs. In two experiments using different tasks, subjects were instructed to respond with their hand or foot following the presentation of an adverb-verb combination. Experiment 1, which prompted reactions via color changes of the stimuli combined with a semantic decision, showed an influence of manner adverbs on response times. This was visible in faster responses following intensifying adverbs compared with attenuating adverbs. Additionally, an interaction between implied verb effector and response effector manifested in faster response times for matching verb-response conditions. Experiment 2, which prompted reactions directly by the adverb type (intensifying vs. attenuating), revealed an interaction between manner adverbs and response effector with faster hand responses following intensifying compared with attenuating adverbs. Additional electroencephalography (EEG) recordings in Experiment 2 revealed reduced beta-desynchronization for congruent verb-response conditions in the case of foot responses along with faster response times. Yet, a direct modulation of verb-motor priming by adverbs was not found. Taken together, our results indicate an influence of manner adverbs on the interplay of language processing and motor behavior. Results are discussed with respect to embodied cognition theories.

**Keywords** Action-related language processing · Manner adverbs · Motor activity · EEG

---

J. Sieksmeyer

Faculty of Natural Sciences, Institute of Experimental Psychology, Heinrich-Heine-Universität, Düsseldorf, Germany

A. Klepp · V. Niccolai (✉) · J. Metzlaff · A. Schnitzler · K. Biermann-Ruben  
Medical Faculty, Institute of Clinical Neuroscience and Medical Psychology,  
Heinrich-Heine-Universität, Düsseldorf, Germany  
e-mail: [valentina.niccolai@hhu.de](mailto:valentina.niccolai@hhu.de)

## 1 Introduction

Embodied cognition theories propose that modal brain regions involved in perception and action are likewise involved in the processing and storage of semantic memory traces (Barsalou, 2008) as opposed to the classical view of amodal brain regions storing semantic memory traces in the form of symbols (for a review see Meteyard, Rodriguez Cuadrado, Bahrami, & Vigliocco, 2012). Specifically, the motor component inherent to language containing action concepts triggers a simulation of the implied movement in sensorimotor areas which is likely reflected in increased activation (Kiefer & Pulvermüller, 2011). The association between perceptual-motor and more cognitive brain areas can be explained through learning experiences, but the precise role of the re-activation of these areas during semantic processing is still under debate (Pulvermüller, 2018). For instance, it is important to elucidate in which detail semantic processing is supported by sensory-motor areas. This issue can be operationalized experimentally in different ways. The current study used adverb-verb combinations to modify the underlying action concept during verbal processing. Behavioral and neurophysiological data were examined in order to find possible interactions of adverbial context with verb processing. These interactions would argue for the contribution of motor cortical areas to language processing being specific and detailed, going beyond superficial epiphenomena.

In turn, theories of language and semantic processing are continuously revised to accommodate empirical findings concerning the functional and neuroanatomical grounding of semantic memory in modality-specific areas (Barsalou, 2008; Binder & Desai, 2011; Pulvermüller, 2018). The current study was aimed at contributing to this discussion by investigating one aspect of embodied cognition, namely the potential interaction between action verb processing and a modifying adverb.

Previous studies reported motor activation either during language processing of action sentences (Aziz-Zadeh, Wilson, Rizzolatti, & Iacobini, 2006; Boulenger, Hauk, & Pulvermüller, 2009; de Vega, Léon, Hernández, Valdés, Padrón, & Ferstl, 2014; Tettamanti et al., 2005) or action verbs (Hauk, Johnsrude, & Pulvermüller, 2004; Kemmerer, Castillo, Talavage, Patterson, & Wiley, 2008; Yang & Shu, 2011). Motor activation seems to be somatotopical in the sense that action verbs implying the movement of a specific extremity elicit activation in corresponding cortical motor areas (Hauk et al., 2004; Pulvermüller, 2005). On the other hand, another study reported no somatotopical activation in primary and premotor areas but rather found action-related activation in the pre-SMA potentially holding an abstract representation of the action verbs in the form of instructional cues (Postle, McMahon, Ashton, Meredith, & de Zubiray, 2008).

On behavioral level, an interaction between action-related language processing and motor execution emerged in altered kinematic measures (Boulenger, Roy, Paulignan, Deprez, Jeannerod, & Nazir, 2006; Dalla Volta, Gianelli, Campione, & Gentilucci, 2009) and in reaction time (Buccino, Riggio, Melli, Binkofski, Gallese, & Rizzolatti, 2005; Sato, Mengarelli, Riggio, Gallese, & Buccino, 2008). Conversely, motor output was shown to affect action language processing (Rüschemeyer,

Lindemann, van Rooij, van Dam, & Bekkering, 2010). Depending on the task and stimulus timing both facilitation (Andres, Finocchiaro, Buiatti, & Piazza, 2015; Glenberg & Kaschak, 2002; Klepp et al., 2017; Scorolli & Borghi, 2007) and interference or prolongation (Boulenger et al., 2006; Klepp, Niccolai, Buccino, Schnitzler, & Biermann-Ruben, 2015; Sato et al., 2008) of motor behavior (i.e. response times) were found.

The demonstrable engagement of motor areas is also evident in studies showing an impairment of action-related language processing in patients suffering from Parkinson's Disease (Fernandino et al., 2013; Herrera, Rodríguez-Ferreiro, & Cuetos, 2011) and Amyotrophic Lateral Sclerosis (Bak, O'Donovan, Xuereb, Boniface, & Hodges, 2001; Grossmann et al., 2008). These impairments indicate an initially important role of motor areas in the ontogenesis of language acquisition (Perniss & Vigliocco, 2014), while neurological disorders affecting motor areas seem to impede an efficient and complete access to semantic memory traces in later life. This can, however, be still partly compensated for by other brain areas (Pulvermüller, 2018). Yet, these results suggest a substantial contribution of motor areas to action-related language processing.

Aside from somatotopy, there is further evidence that sensorimotor involvement in language processing is specific and detailed. For instance, it may reflect semantic features of verbal material: The amount of effector-specific movement affected verb-motor priming (Klepp et al., 2017). Additionally, functional magnetic resonance imaging (*fMRI*) activity in parietal areas within the motor network was modulated by the specificity of action plans described by verbs (van Dam, Rüschemeyer, & Bekkering, 2010). Activity in pre-motor areas can reflect motor features described in action sentences, e.g. the degree of physical effort the described action requires as determined by a verb-object combination (Moody & Gennari, 2010).

In natural language, however, important cues about the precise implied action may not only come from the verb itself, but from other sources such as objects and adverbial constructions. The linguistic focus hypothesis (*LFH*) postulated by Taylor and Zwaan (2008) suggests that motor simulation, which is assumed in the theoretical framework of embodied cognition theories regarding action-related language processing, is dependent on the linguistic focus. If the described action is maintained within the linguistic focus, e.g. through action-modifying adverbs, motor simulation of the action occurs beyond the action verb itself and continuation of motor activity should be observed; if the linguistic focus is shifted away from the action, e.g. through agent-modifying adverbs, no motor simulation occurs and termination of motor activity should be observed. The study conducted by Taylor and Zwaan (2008) demonstrated that adverbs can influence reading times. Participants were instructed to read a sentence frame by frame by turning a knob either in clockwise or counter-clockwise direction. The sentences contained hand action verbs depicting either clockwise or anticlockwise movements followed by an adverb either modifying the action (e.g. *quickly*) or the agent (e.g. *happily*). Facilitation of reading times occurred in direction-matching versus non-matching verb-response conditions and when adverbs modified the action instead of the agent.

Yet it remains unclear, if the motor simulation of the action verb is differentially affected by different kinds of adverbs. For example, despite relating to a bodily action, action verbs might not contain lexically specific information about the amount of force with which the action is executed (Goldschmidt, Gamerschlag, Petersen, Gabrovska, & Geuder, 2017). This questionnaire study examined whether the force component in the German action verb “schlagen” (*to hit*) could be modulated in sentences containing adverbs. Crucially, force-denoting manner adverbs (*lightly/hard*) directly modified the action’s force component in the direction suggested by the adverb. Moreover, force modification may also be achieved by inferences through agent-oriented adverbs (Goldschmidt et al., 2017). The current study uses manner adverbs denoting a clear attenuation or intensification of either the force component or the speed component, thus expected to directly modify the action described by verbs. Note that the term “force” is used as a synonymous expression for “intensity” and is not used in terms of causality, as it is in the linguistic field of “force semantics” and “force dynamics”.

Based on these findings we investigated in two separate experiments if adverbs further influence reaction times in a well-established priming paradigm containing hand/foot action verbs and hand/foot responses (Klepp et al., 2017). In both experiments, the interaction of verb type and response effector was anticipated as a priming effect resulting in faster response times for congruent verb-response conditions. We furthermore introduced intensifying and attenuating manner adverbs as an additional factor. The previously observed interaction of verb type and response effector was hypothesized to be more pronounced when the action verb was combined with an intensifying compared to an attenuating adverb resulting in even faster response times in congruent verb-response conditions. Adverb-verb (Experiment 1) and verb-adverb (Experiment 2) order of presentation realized in two separate experiments additionally allowed investigating the influence of the time point of adverb presentation with respect to the action verb processing stream.

A useful technique to investigate the time course of activation with respect to action verb processing is EEG. Increased activity in motor areas is typically accompanied by increased desynchronization in the mu (10–15 Hz) and the beta band (15–25 Hz). This oscillatory pattern is generally associated with motor preparation and execution (Pfurtscheller & Lopes da Silva, 1999; Pfurtscheller, Neuper, Andrew, & Edlinger, 1997). Typically, desynchronization increases, reaching a peak during response execution, while a rebound consisting in increased synchronization is found about a second after movement offset (Pfurtscheller, & Lopes da Silva, 1999). A similar pattern has also been observed during action verb processing (van Elk, van Schie, Zwaan, & Bekkering, 2010; Moreno, de Vega, Léon, Bastiaansen, Lewis, & Magyari, 2015; Niccolai, Klepp, Weissler, Hoogenboom, Schnitzler, & Biermann-Rubén, 2014). Neural oscillatory and event-related potential (ERP) effects related to the presentation of action verbs have been reported as early as 170–250 ms after word onset (Pulvermüller, Härle, & Hummel, 2000, 2001; van Elk et al., 2010), displaying somatotopy (Hauk et al., 2004; Pulvermüller et al., 2001). This pattern of results suggests that motor areas contribute to the processing of verbal linguistic action stimuli.

We therefore conducted EEG measurements in Experiment 2 to gain further insights in the processing of the action verb and its possible interaction with the adverb. According to the somatotopic organization of the motor cortex, differential effects for hand and foot responses were hypothesized in electrode sites C3 and Cz, respectively. We expected stronger hand-related activity at electrode site C3 and stronger foot-related activity at electrode site Cz. This approach has been successfully used regarding evoked EEG activity during action verb reading before (Hauk & Pulvermüller, 2004; Pulvermüller et al., 2000, 2001). We specifically focused on the mu and beta band in our study due to their role in motor processes (Pfurtscheller & Lopes da Silva, 1999) and action-related language processing (Klepp et al., 2015, Moreno et al., 2015; van Elk et al., 2010). Thus, we expected the mu and beta desynchronization around the onset of the response to be reduced in congruent conditions due to the priming effect of the action verbs (Grisoni, Dreyer, & Pulvermüller, 2016; Schacter, Wig, & Stevens, 2007).

As the manner adverb directly modified the action verb which elicits motor activity if semantically processed, we expected the adverb to further modulate the language-motor interaction. The interaction of verb type and response effector was hypothesized to be more pronounced when the action verb was combined with an intensifying compared to an attenuating adverb, resulting in reduced mu and beta desynchronization in congruent verb-response conditions.

Methods and Results of Experiments 1 and 2 will be reported separately followed by a joint discussion.

## 2 Experiment 1

### 2.1 Methods

#### 2.1.1 Participants

Thirty-two participants (eleven male; mean age = 24.97 years,  $SD = 6.71$ ) were included into the study. Exclusion criteria were academic linguistic expertise, history of prior neurological or psychiatric disorders and medication affecting the central nervous system. Participants were monolingual German native speakers with normal or corrected-to-normal vision. Right-handedness and footedness was assessed and confirmed. Hand dominance was assessed with the Hand Dominance Test (*HDT*, Steingrüber, 2011), as well as with the German version of the Edinburgh Handedness Questionnaire (*EHI*, Oldfield, 1971). Right footedness was tested with a self-report questionnaire extracted from the Lateral Preference Inventory (*LPI*, Ehrenstein & Arnold-Schulz-Gahmen, 1997).

This experiment is in accordance with the Declaration of Helsinki and was approved by the ethics committee of the Medical Faculty at Heinrich-Heine-University Düsseldorf (study number: 3400). All subjects gave written informed

consent before the beginning of the experiment and received course credit or financial reimbursement.

Two subjects were excluded due to data loss. Three subjects exceeded our criterion of at most 10% incorrect responses during the experiment. After the experimental session, one participant reported taking medication affecting the central nervous system and was also excluded. The final set of participants in Experiment 1 therefore consisted of 26 subjects (nine male, mean age = 25.28 years,  $SD = 7.39$ ).

### 2.1.2 Stimuli

We used a total of 36 disyllabic German verbs and eight German adverbs (for an overview see Table 1). The verb set consisted of three categories with 12 verbs each: manual actions, e.g. “klatschen” (*to clap*), foot actions, e.g. “rennen” (*to run*) and abstract actions, e.g. “denken” (*to think*). We used a subset of verbs out of a previous selection which had been based on successive rating and matching procedures (details see Klepp et al., 2014) including familiarity, imageability and movement energy, as well as word length and verb frequency (Leipzig Corpora Collection, *LCC*, Biemann, Heyer, Quasthoff, & Richter, 2007, available at <http://wortschatz.uni-leipzig.de>). As indicated by a multivariate ANOVA with verb category as an independent variable the final set of verbs differed with regard to some of these variables, but only in comparison to the abstract verb category. This category served as the No Go-condition, however and was not further analyzed. Importantly, hand and foot verbs did not differ significantly (all  $p > 0.087$ ).

Twenty-five adverbs entered a rating process ( $n = 4$ ) serving to redefine stimuli selection. Participants were asked to evaluate the probability of verb and adverb going together (from “not at all” to “absolutely”). This was termed as the semantic fit of adverb-verb combinations. Adverb selection was based on their semantic fit with the previously selected set of action verbs, as well as the possibility to define opposed pairs of intensifying and attenuating adverbs, e.g. “kräftig”—“kraftlos” (*strongly—feebly*). This resulted in the subsequent inclusion of eight out of the initially selected 25 adverbs, four of which strengthening or accelerating the movement implied by the verb (intensifying adverbs) and four weakening or slowing the implied action (attenuating adverbs). Exact Mann-Whitney-*U*-Test revealed no significant differences in the frequency of intensifying and attenuating adverbs ( $U = 14.50, p = 0.343$ ) nor differences in their semantic fit to hand verbs ( $U = 6.50, p = 0.645$ ) or foot verbs ( $U = 5.00, p = 0.381$ ), respectively. Please note that linguistically all adverbs are adjectives applied in an adverbial manner of use. For the sake of simplification we will refer to them as “manner adverbs” in this article.

### 2.1.3 Procedure

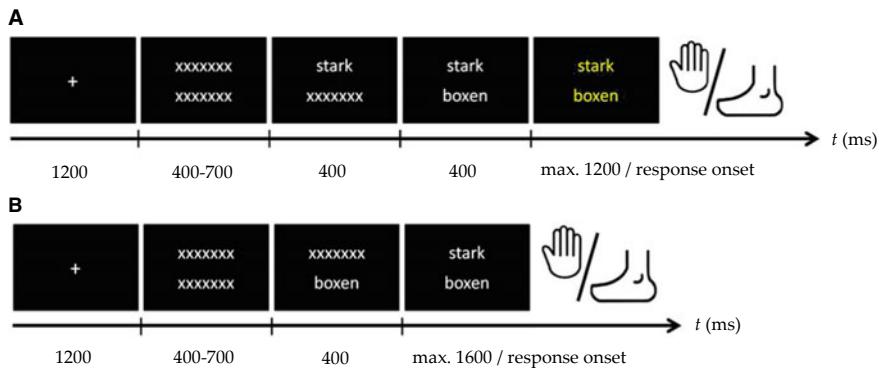
Subjects were seated at a distance of 95 cm from a computer screen (ASUS VG248, ASUS Computer International, Fremont, California, USA) with a keyboard in front

**Table 1** The stimulus set consisted of 36 verbs and 8 adverbs. All adverbs and verbs are reported in German and English, respectively, although only the German version was used in the study

Verb category	Hand	Foot	Abstract
	boxen ( <i>to box</i> )	hinken ( <i>to limp</i> )	achten ( <i>to respect</i> )
	fuchteln ( <i>to flail</i> )	humpeln ( <i>to hobble</i> )	ahnen ( <i>to guess</i> )
	klatschen ( <i>to clap</i> )	kicken ( <i>to kick</i> )	büßen ( <i>to atone</i> )
	kneten ( <i>to knead</i> )	laufen ( <i>to walk</i> )	denken ( <i>to think</i> )
	paddeln ( <i>to paddle</i> )	rennen ( <i>to run</i> )	gönnen ( <i>to indulge</i> )
	rubbeln ( <i>to rub</i> )	schlittern ( <i>to slide</i> )	grübeln ( <i>to ponder</i> )
	scheuern ( <i>to scorn</i> )	schlurfen ( <i>to shuffle</i> )	hoffen ( <i>to hope</i> )
	schrubben ( <i>to scrub</i> )	stampfen ( <i>to stamp</i> )	merken ( <i>to remember</i> )
	stupsen ( <i>to nudge</i> )	strampeln ( <i>to struggle</i> )	träumen ( <i>to dream</i> )
	tippen ( <i>to type</i> )	tänzeln ( <i>to prance</i> )	wünschen ( <i>to wish</i> )
	trommeln ( <i>to drum</i> )	trampeln ( <i>to tramp</i> )	wundern ( <i>to marvel</i> )
	zupfen ( <i>to pick</i> )	treten ( <i>to tread</i> )	zweifeln ( <i>to doubt</i> )

Adverb category	Intensifying	Attenuating
	kräftig ( <i>forcefully</i> )	kraftlos ( <i>feebley</i> )
	hektisch ( <i>hectically</i> )	ruhig ( <i>calmly</i> )
	stark ( <i>strongly</i> )	zaghaft ( <i>tentatively</i> )
	flink ( <i>swiftly</i> )	träge ( <i>duelly</i> )

of them and a foot pedal (USB Triple Foot Switch II; Scythe, Tokyo, Japan) positioned under the table. All trials started with a black background screen containing a white fixation cross at the center presented for 1200 ms. This was followed by the presentation of a mask consisting of two horizontal lines of seven white 'X' for a jittered interval between 400 and 700 ms. Thereafter, an adverb was presented in white letters pseudorandomly above or below the fixation cross for 400 ms together with the remaining upper or lower line of seven 'X'. The verb followed in white letters replacing the latter seven 'X' and the adverb-verb combination was displayed together for another 400 ms. Then, the stimuli turned either blue or yellow. Subjects were instructed to respond as fast and accurately as possible with the hand or the foot according to the color of the adverb-verb combination, but only if the verb expressed



**Fig. 1** Experimental procedure. **A** Experiment 1. **B** Experiment 2

a concrete bodily action. Participants were pseudorandomly assigned to one of two groups of color change instructions: color change to blue required a hand response by pressing the ‘space’-key on a keyboard while a color change to yellow required a foot response by pressing down a foot pedal for 50% of the subjects. For the other 50%, the assignment was reversed. The experimental procedure of each trial is shown in Fig. 1A.

Pseudorandomization of the spatial position of verb and adverb was introduced to prevent participants from adopting a strategy to solely attend to the stimulus relevant for solving the experimental task, i.e. the verb in Experiment 1 and the adverb in Experiment 2, respectively. The spatial predictability of these stimuli could have resulted in impaired semantic processing of adverb-verb-combinations, which we tried to preclude by variation of the spatial positions.

Each adverb was combined with each verb and presented once with each type of color change resulting in a total of 576 trials per subject. The experiment lasted about 50 min. Stimuli were presented using Presentation 14.9 software (Neurobehavioral Systems, Albany, California, USA).

#### 2.1.4 Statistical Analysis

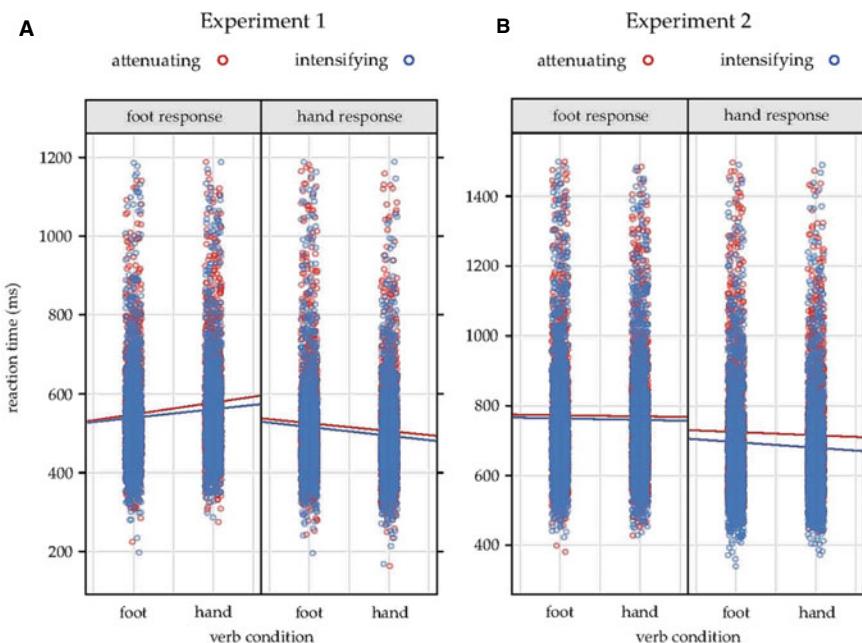
We computed a linear mixed effect model using the package lme4 (version 1.1–13, Bates, Maechler, Bolker, & Walker, 2015) for R (version 3.3.3) including crossed random effects for subjects and items. This method is especially advantageous for studies incorporating psycholinguistic stimuli since it is assumed that not only participants but also the items are randomly drawn from a population (Baayen, Davidson, & Bates, 2008). Linear mixed effect models allowed the inclusion of the two-level factor verb (hand, foot), the two-level factor adverb type (intensifying, attenuating) and the two-level factor response effector (hand response, foot response). Thus, the fixed effects included the factors verb, adverb type and response effector, as well as their two-way and three-way interactions. Random effects for

participants included random intercepts with random slopes for the factors verb, adverb type and response effector. Random effects for items only included random intercepts. All analyses used logarithmically transformed reaction times of correct responses within 150 to 1500 ms.  $T$ -values below  $-2$  or above  $2$  are considered to represent significant effects. Post hoc tests were calculated using the package lsmeans (version 2.25–5, Lenth, 2016).

## 2.2 Results

### 2.2.1 Behavioral Data

Errors and responses faster than 150 ms or slower than 1500 ms after the Go-signal onset were excluded resulting in the exclusion of 358 trials (3.59%). The Go-signal is defined as the cue stimulus prompting a response, i.e. here the color change of the adverb-verb-combination. Raw data are shown in Fig. 2A. We observed a significant main effect of response effector ( $t = 4.86$ ) with faster hand responses than foot

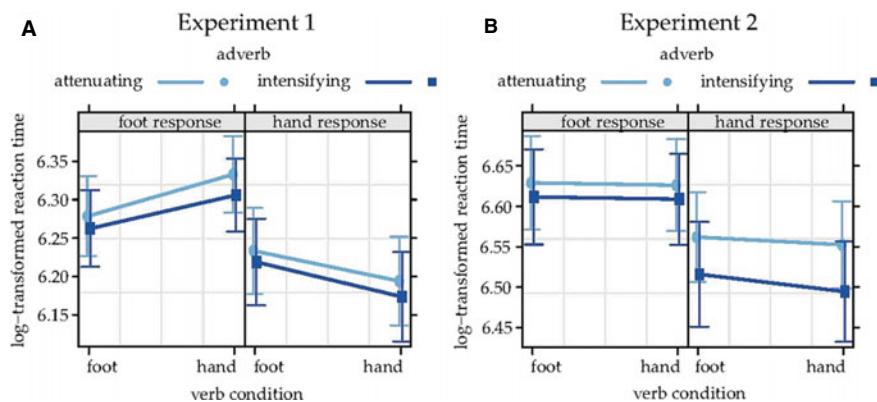


**Fig. 2** Data distribution in Experiment 1 (A) and Experiment 2 (B). Raw data is split for verb conditions (hand verb, foot verb) on the x-axis, y-axis indicates the response times in milliseconds. Data is furthermore split according to the response effector (hand responses, foot responses) following intensifying (blue) and attenuating (red) adverbs. Red and blue lines indicate mean values in the respective conditions

responses. The main effect of adverb type was significant ( $t = 4.01$ ) as well, with faster responses following intensifying adverbs compared to attenuating adverbs. Furthermore, the interaction between verb and response effector was significant ( $t = -11.20$ ). Post hoc tests indicated significantly faster ( $z = 3.25, p = 0.001$ ) hand responses following hand verbs compared to foot verbs and significantly faster ( $z = -3.76, p < 0.001$ ) foot responses following foot verbs compared to hand verbs. The hypothesized three-way interaction of verb, response effector and adverb type was not significant ( $t = -0.32$ ). All model estimates are given in Table 2. Fitted model parameters ( $\pm SD$ ) for verb x adverb type x response effector are depicted in Fig. 3A.

**Table 2** Results of statistical analyses of behavioral data. Model estimates ( $\beta$ ), standard error (SE) and  $t$ -values are reported for Experiment 1 (left) and Experiment 2 (right). Significant effects are bold

	Experiment 1			Experiment 2		
	$\beta$	SE	$t$	$\beta$	SE	$t$
Verb	-0.002	0.006	-0.27	0.005	0.003	1.63
Adverb type	<b>0.010</b>	<b>0.002</b>	<b>4.01</b>	0.017	0.010	1.66
Response effector	<b>0.045</b>	<b>0.009</b>	<b>4.86</b>	<b>0.044</b>	<b>0.008</b>	<b>5.65</b>
Verb x adverb type	-0.002	0.002	-1.03	-0.001	0.002	-0.79
Verb x response effector	<b>-0.023</b>	<b>0.002</b>	<b>-11.20</b>	-0.003	0.002	-1.70
Adverb type x response effector	0.001	0.002	0.47	<b>-0.009</b>	<b>0.002</b>	<b>-4.68</b>
Verb x adverb type x response effector	-0.001	0.002	-0.32	0.002	0.002	0.82



**Fig. 3** Effects of Experiment 1 (A) and Experiment 2 (B). The 2-level factor verb (hand verb, foot verb) is denoted on the x-axis, y-axis indicates log-transformed response times. Data are split according to adverb type (intensifying adverb, attenuating adverb) and response effector. Circles, squares and error bars indicate fitted model parameters with standard deviation

### 3 Experiment 2

#### 3.1 Methods

##### 3.1.1 Participants

Seventeen participants (six male; mean age = 24.82 years,  $SD = 4.90$ ) were included into the study. Exclusion criteria were the same as in Experiment 1. Handedness and footedness was assessed as in Experiment 1. The experiment is in accordance with the Declaration of Helsinki and was approved by the ethics committee of the Medical Faculty at Heinrich-Heine-University Düsseldorf (study number: 3400). All subjects gave written informed consent before the beginning of the experiment and received course credit or financial reimbursement.

Two subjects were excluded due to unidentifiable artifacts in the EEG recordings and excessive eye blinks during stimulus presentation. The final set of participants in Experiment 2 consisted of 15 subjects (six male, mean age = 24.93 years,  $SD = 5.23$ ).

##### 3.1.2 Stimuli

The stimulus set included the same adverbs and concrete verbs as in Experiment 1, i.e. 24 concrete verbs (twelve hand verbs and twelve foot verbs) and eight adverbs (four intensifying adverbs, four attenuating adverbs).

##### 3.1.3 Procedure

All subjects participated in two separate experimental sessions at least seven days apart. Sessions differed regarding the instructions: in one session subjects had to react with their right hand to intensifying adverbs and with their right foot to attenuating adverbs. Reaction times were recorded. The response effector-adverb type relationship was reversed in the other session. The order of sessions was counterbalanced across subjects. Experimental sessions were conducted in an electrically shielded room. The experimental setup was the same as described in Experiment 1 with only few changes. First, the verb preceded the adverb. Second, the onset of the adverb instantaneously cued the response effector according to the respective instructions of the current session, i.e. there was no color change. Third, the verb-adverb combination was presented until response onset or maximally 1600 ms. The trial design is depicted in Fig. 1B. Each adverb was paired with each verb and each combination was shown twice, thus resulting in a total of 384 trials per subject. Each experimental session lasted about 30 min. Stimuli were presented using Presentation 14.9 software (Neurobehavioral Systems, Albany, California, USA) in white font on a black background.

### 3.1.4 EEG Data Acquisition

We recorded the EEG signal with 29 Ag–AgCl electrodes mounted in an elastic cap (EASYCAP GmbH, Herrsching, Germany), according to the 10/20 system. The average of right and left mastoid was used as reference and electrode position AFz as ground. Vertical EOG was recorded using bipolar electrodes. EEG-signals were amplified with a BrainAmp MR Plus amplifier (Brain Products, Munich, Germany). A sampling rate of 1000 Hz and an online high-pass filter of 0.3 Hz were applied. Impedance of all electrodes was kept below 10 kΩ. EEG and EOG signals were registered with BrainVision Recorder (Brain Products GmBH, Munich, Germany).

### 3.1.5 EEG Data Processing

Neurophysiological data were analyzed with Fieldtrip (version 20160629, Oostenveld, Fries, Maris, & Schoffelen, 2011), an open source toolbox for Matlab (version R2016a, Mathworks, Natick, Massachusetts, USA). Episodes include time-windows from 2 s before verb onset to 0.5 s after response onset. A semi-automatic artifact detection routine was applied to identify electrode jumps and muscle artifacts. A lowpass filter at 120 Hz was applied and line noise at 50 and 100 Hz filtered out. Trials were visually inspected for blink artifacts in the critical time window between verb and response onset as well as for non-EOG artifacts in the whole trial. Trials containing artifacts were rejected. Remaining blink artifacts in the baseline or post-response period were removed using independent component analysis (ICA).

Data were subsequently split into eight conditions defined by adverb type (intensifying vs. attenuating), verb (hand vs. foot) and response effector (hand vs. foot) and entered a time-frequency analysis. To discern and investigate semantic processing around adverb onset and motor processes during response execution more closely, we conducted two analyses locked to adverb and response onset, respectively. In both analyses data were aligned to the respective event, with 0 either denoting the onset of adverb or response. Time-frequency representations (*TFRs*) were computed in steps of 2 Hz from 2 to 30 Hz using a Fourier transformation. We applied a single Hanning taper with a width of 5 cycles, sliding in steps of 40 ms. Data were baseline-corrected using a time window of  $t_a = -1.3$  to  $-0.8$  s for the *adverb-locked* analysis and  $t_r = -1.5$  to  $-1.0$  s for the *response-locked* analysis.

### 3.1.6 Statistical Analysis

#### Behavioral Data

The linear mixed effect model contained the two-level factors verb (hand, foot), adverb type (intensifying, attenuating) and response effector (hand response, foot response). Fixed effects included the factors verb, adverb and response and their interactions. Random effects for participants included random intercepts for subjects

and random slopes for the factors verb, adverb and response. Random effects for items only included random intercepts. Logarithmically transformed reaction times of correct responses within 150 to 1500 ms entered the analysis.  $T$ -values below  $-2$  or above  $2$  are considered to represent significant effects. Post hoc tests were carried out using the package lsmeans (version 2.25-5, Lenth, 2016).

### EEG Data

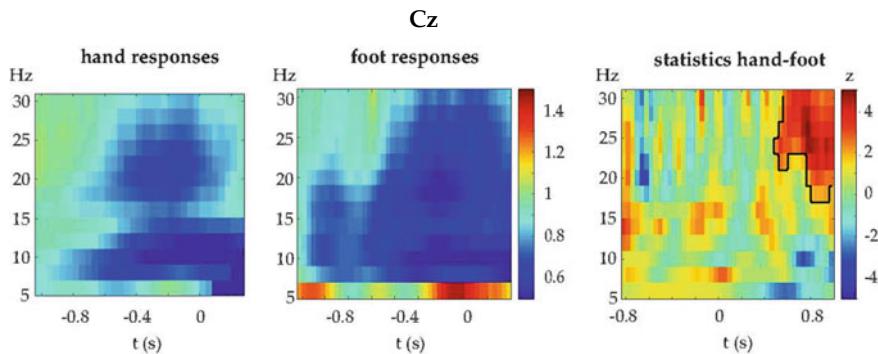
To statistically analyze the EEG data we computed pseudo- $t$ -values for each participant to normalize individual differences (compare Lange, Oostenveld, & Fries, 2011). These  $t$ -values were then transformed into  $z$ -values to account for different number of trials in each condition (Klepp et al., 2015; see van Dijk, Nieuwenhuis, & Jensen, 2010). Then we applied a non-parametric statistical procedure to assess significant differences on the group level. This non-parametric randomization approach identifies clusters containing neighboring timepoints and frequencies while simultaneously correcting for multiple comparisons (Maris & Oostenveld, 2007).

Conditions were considered significantly different if the test statistic obtained from 5000 permutations resulted in an alpha-level below 0.05. We defined the relevant contrasts of conditions based on the resulting significant behavioral effects. In addition, we investigated if semantic priming is mirrored in reduced desynchronization in congruent verb-response conditions, as stated in our hypothesis. Electrodes C3 as proxy of the right hand and Cz as proxy of the right foot were analyzed separately.

## 3.2 Results

### 3.2.1 Behavioral Data

Errors and responses faster than 150 ms or slower than 1500 ms after adverb onset were excluded from further analysis resulting in the exclusion of 411 trials (3.57%). Raw data are shown in Fig. 2B. The mixed model analysis showed a significant main effect for response effector ( $t = 5.65$ ) with faster hand responses than foot responses. A significant interaction between the factors adverb type and response effector emerged ( $t = -4.68$ ). Post hoc tests revealed significant differences for hand responses ( $z = 2.448, p = 0.014$ ) with faster hand responses following intensifying compared with attenuating adverbs. No difference emerged in the case of foot responses ( $z = 0.820, p = 0.412$ ). The interaction between verb and response effector was not significant ( $t = -1.70$ ). The three-way interaction of verb, response effector and adverb type was not significant ( $t = 0.82$ ). All values are given in Table 2. Fitted model parameters ( $\pm SD$ ) for verb x adverb type x response effector are depicted in Fig. 3B.



**Fig. 4** Grandaverage EEG data and statistical contrasts relating to significant effects in the adverb-locked analysis of Experiment 2. The time in seconds is depicted on the x-axis with 0 denoting the onset of adverb, the frequency in Hertz is shown on the y-axis. Data is furthermore color-coded according to the power relative to baseline (left and middle column) or according to the z-value (right column) of the respective statistical comparison. The contrast shows hand- and foot response-related activity in electrode Cz with the significant cluster outlined in black.

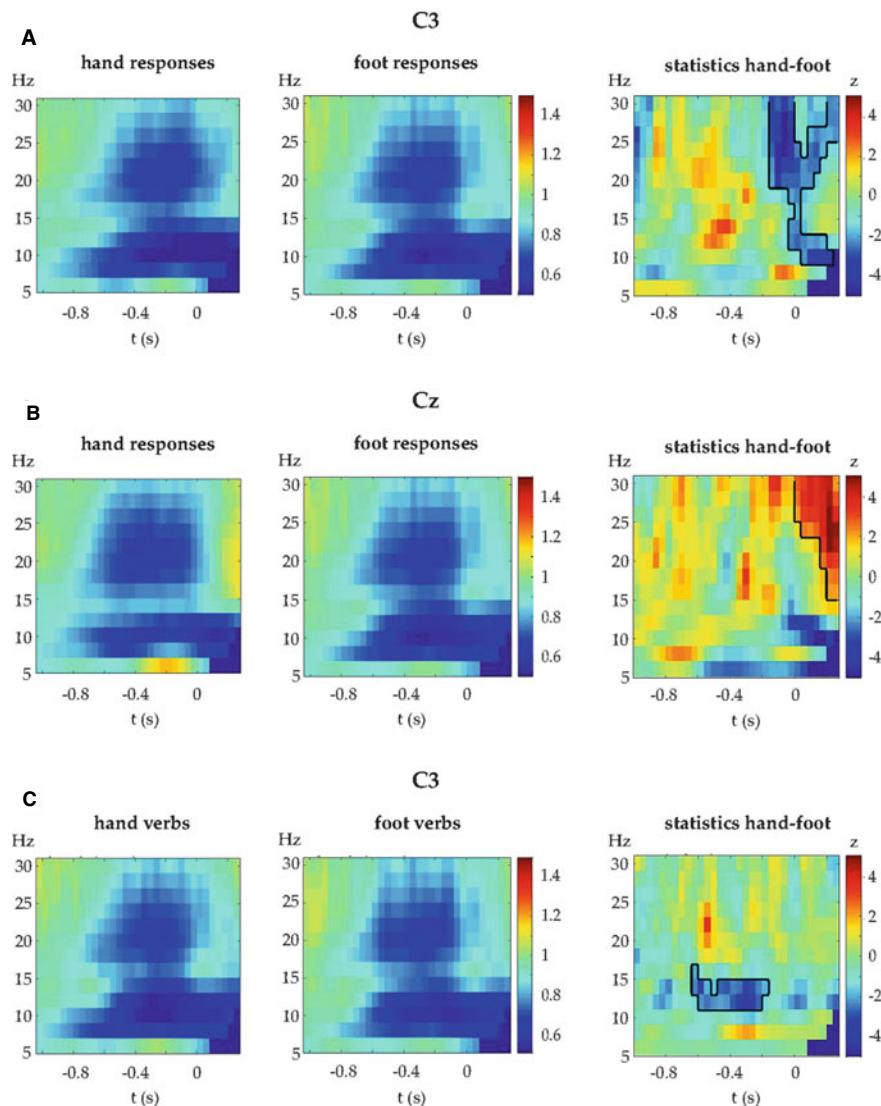
### 3.2.2 EEG Data

#### Adverb-Locked Analysis

A cluster between  $t_a = 0.48$  and 1 s after adverb onset ( $p = 0.003$ ) indicated a significant difference between response effectors in the electrode Cz ranging from 17 to 30 Hz, i.e. stronger beta desynchronization for foot responses (Fig. 4). No effect was found in C3 (all  $p > 0.201$ ). Neither adverbs nor any interaction with response effector or verb showed significant effects in electrodes C3 nor Cz (all  $p > 0.110$ ).

#### Response-Locked Analysis

A cluster at  $t_r = -0.16$  to 0.32 s ( $p < 0.001$ ) indicated a significant difference between hand and foot responses in electrode C3 ranging from 9 to 30 Hz showing stronger mu and beta desynchronization for hand responses (Fig. 5A). Complementarily, in electrode Cz, a cluster indicated a significant difference between the hand and foot condition at  $t_r = 0$ –0.4 s after response onset ( $p = 0.002$ ) ranging from 15 to 30 Hz, i.e. stronger beta desynchronization for foot responses (Fig. 5B). A significant cluster ( $p = 0.019$ ) in electrode C3 showed that in the case of foot responses, the hand verb condition showed significantly more beta desynchronization than the foot verb condition at  $t_r = -0.64$  to 0.16 s ranging from 12 to 18 Hz (Fig. 5C). No effect was observed in electrode Cz ( $p > 0.082$ ). No corresponding effect emerged for hand responses (all  $p > 0.116$ ).



**Fig. 5** Grandaverage EEG data and statistical contrasts relating to significant effects in the response-locked analysis of Experiment 2. The time in s is depicted on the x-axis with 0 denoting the onset of response, the frequency in Hz is shown on the y-axis. Data is furthermore color-coded according to the power relative to baseline (left and middle column) or according to the z-value (right column) of the respective statistical comparison. **A:** Hand response- and foot response-related activity in electrode C3 with the significant cluster outlined in black. **B:** Hand response- and foot response-related activity in electrode Cz with the significant cluster outlined in black. **C:** Foot response-related activity following hand verbs and foot verbs with the significant cluster outlined in black.

### 3.3 Discussion

Our results show an influence of manner adverbs on motor behavior. In Experiment 1, we found a significant main effect of adverb indicating faster responses following intensifying compared with attenuating adverbs. This effect might depend on the direct relation between the force component of the action verb and the manner adverb specifying the amount of force implied in the movement (Goldschmidt et al., 2017). Action verbs are reported to elicit motor activation (Hauk et al., 2004; Pulvermüller, 2005), especially when processed semantically (Klepp et al., 2017; Sato et al., 2008). Motor output interacts with action-related language processing because of shared neuronal circuits (Boulenger et al., 2006; Dalla Volta et al., 2009). Manner adverbs modifying an action verb might therefore modulate its elicited motor activation by modulating the amount of force implied in the action. As was shown in the case of imageability (Klepp et al., 2015) and effector-specific movement (Klepp et al., 2017), semantic features of the action verb might influence motor behavior. Cortical motor areas might therefore also be involved in the processing of semantic features of stimuli relating to the action verb. This is furthermore corroborated by the complementary results of Experiment 2. Here, though no significant main effect of adverb type emerged, manner adverbs interacted with the response effector. Hand responses following intensifying adverbs were significantly faster than hand responses following attenuating adverbs. Participants had to respond depending on the adverb type.

Comparing Experiments 1 and 2 it seems that the main effect of adverb type in Experiment 1 switched to an interaction between adverb type and response effector in Experiment 2. The main difference between these two experiments is the order of adverb-verb (Experiment 1) and verb-adverb (Experiment 2) presentation combined with the instruction cues *color change* (Experiment 1) and *adverb type* (Experiment 2). The relevance and subsequent psycho-linguistic processing of the adverb for successfully operating on the tasks hence was different in Experiments 1 and 2: In Experiment 1, priming of force components could have taken place resulting in a main effect of adverb type even though the semantics of the adverbs were of minor relevance. In Experiment 2 on the other hand, the semantics of the adverb were indicative for the required response potentially resulting in simulation processes directly interacting with response preparation. In addition, while both tasks prompted semantic processing of the verbal material, Experiment 2 might have increased the participants' awareness of the semantic features of the manner by making them task-relevant. Studies concerning the mental timeline argued that mental simulations only occur during language processing, if the semantic features of the verbal material is task-relevant and the processor is aware of these features (Maienborn, Alex-Ruf, Eikmeier, & Ulrich, 2015; Ulrich & Maienborn, 2010). The increased awareness of the semantic features may have resulted in the more specific interaction between adverb and response. That the effect was only found for hand responses and not for foot responses might be attributable to the closer connection of the hand with language (Rizzolatti & Arbib, 1998). Another explanation could be that due to

longer response times for foot responses the interaction with adverbs might fade with processing time. Still, foot responses to intensifying adverbs were numerically faster than to attenuating adverbs. Arguably, intensifying adverbs might increase motor activation per se thereby increasing the motor contribution to the processing of the action verb. This might reflect a semantic priming effect even for manner adverbs. This remains elusive, however, since no corresponding effect of manner adverbs was found in the neurophysiological data of Experiment 2. Differential motor activation relating to intensifying and attenuating adverbs could arise in studies focusing solely on the semantic processing of manner adverbs. Yet both experiments reported in this study incorporated manner adverb-action verb combinations which might have limited our ability to discern adverb- and verb-related processes regarding brain oscillations.

In addition to the effects of manner adverbs, action verbs interacted with motor behavior. In Experiment 1, facilitated hand and foot responses in congruent compared with incongruent verb-response combinations revealed a semantic priming effect, which is in line with previous findings (Scorolli & Borghi, 2007; Klepp et al., 2017). This is corroborated by the neurophysiological data recorded in Experiment 2. Results showed reduced beta desynchronization in electrode C3 for foot responses following foot verbs compared to hand verbs. As expected, the congruent condition presented with reduced motor activation (Grisoni et al., 2016; Schacter et al., 2007). The onset of the effect was about 600 ms before response onset, which would have allowed the action verb to be processed semantically and subsequently interacting with response execution (Kutas & Hillyard, 1984). However, the effect emerged in electrode C3 only, which was located approximately above the cortical motor hand area, whereas no complementary effect for hand responses emerged in electrode Cz. Furthermore, no verb x response effector interaction was visible in the behavioral data of Experiment 2. Hence, the observed differences might alternatively be accounted for by significant differences between hand and foot verbs, respectively. This might be due to the limited set of action verbs employed in this study. To reduce confounding effects of imageability, familiarity and frequency, we matched our verb set very carefully. However, this might have prevented us from mapping a wider range of possible differences in the action verbs, for instance with regard to their movement pattern as well as other linguistic features. Differences in brain oscillations for foot responses following hand and foot verbs might therefore alternatively be unrelated to a semantic priming effect but merely reflect an overall stronger desynchronization following hand verbs compared with foot verbs independent of the response effector.

Two further aspects should be discussed, namely the somatotopy of response effectors and timing. Hand and foot responses were required in a double-dissociation paradigm. As visible in the behavioral data of both experiments, hand responses were overall faster than foot responses, which is in line with previous findings (Buccino et al., 2005; Gianelli & Dalla Volta, 2015; Klepp et al., 2017). Experiment 2 indicated differential motor activation for hand and foot responses in electrodes C3 and Cz in analyses around adverb and response onset. Oscillatory differences arose predominantly in the beta frequency range in a time window relating to response execution. Crucially, stronger desynchronization for hand responses was observed

in electrode C3, while stronger desynchronization for foot responses emerged in electrode Cz. Our results thus demonstrate somatotopical activity differences related to the respective response effector, as hypothesized. This should have allowed for the detection of differential EEG effects of verbs and adverbs for the two response effectors. Indeed, stronger desynchronization for hand than foot verbs preceding foot responses in electrode C3 was found, but not the full pattern of effects expected from the double dissociation paradigm. One straightforward explanation may be that when no behavioral effects were found, there simply might have been no differences in neurophysiological processing to be measured by EEG. Note, however, that neurophysiological effects are sometimes reported in the absence of behavioral differences (Mollo, Pulvermüller, & Hauk, 2016). Nevertheless, the paradigm of Experiment 2 may be not optimally suited for the detection of language-motor priming effects. More specifically, the temporal proximity between manner adverb and hand/foot responses onset in Experiment 2 might have been too close to discern oscillatory differences in the semantic processing of manner adverbs. Instead, potentially subtle activity differences relating to the processing of the manner adverb might have been overshadowed by motor activation induced during response execution processes. In addition, potential activation differences might have been observable in other electrodes, especially located above other language-related brain areas, e.g. temporal regions; these regions might also reflect differences based on the type of manner adverb or its interaction with action verbs.

An important concern in the comparison between the effects of Experiment 1 and Experiment 2 is the temporal structure of stimulus presentation and the average response time. In Experiment 1, the average time interval between adverb and response onset was 1300 ms while the average time interval between action verb and response, on the other hand, was only 900 ms. There was a priming effect of verb effector, but only an unspecific effect of adverb type. Thus, Experiment 2 was designed to induce more semantic interaction with the hypothesis to find an interaction of priming and adverb type, reflected in neurophysiological data. The reversal of verb and adverb presentation order also implied that the average time interval between verb presentation and response was 1150 ms, with 750 ms between adverb and response onset. Action verbs did not influence response times possibly due to the prolonged interval between verb and response onset. Accordingly, stimulus-response intervals likely modulated the effects observed in the two experiments.

Further, a relatively small sample size and the inclusion of only two EEG electrodes in the statistical analyses might have limited the power of our results. Inclusion of a greater sample to increase effect sizes and a greater number of EEG electrodes could lead to a more detailed picture regarding the interplay of action-related language processing and motor activity and its modulation by manner adverbs.

Future studies should furthermore investigate which semantic aspects of manner adverbs potentially elicit motor activation, e.g. differentiating between force and velocity, providing closer insights into the extent of motor involvement in language processing. The small number of force- and velocity-modulating adverbs (two each for intensifying and attenuating adverbs) prevented us from validly deducing differential effects on the verb-response interaction. On the other hand, some action verbs,

e.g. “boxen” (*to box*), might be predominately modulated by adverbs defining the amount of force, while others, e.g. “tippen” (*to type*), might be more susceptible to an adverbial modulation relating to the velocity of the action. This raises the important question whether such differences are mirrored in the overt motor behavior or neurophysiological activity. Additionally, previous studies suggested an influence of various movement-dependent factors on beta-desynchronization in motor areas (Tan et al., 2013; Nakayashiki, Saeki, Takata, Hayashi, & Kondo, 2014). A consecutive study might therefore also be concerned with the influence of manner adverb on the motor response by taking various movement-related parameters into consideration. Furthermore, adverb-verb combinations should be included in natural sentences to shed more light on the influence of grammatical constructions on action-related language processing in sensorimotor areas.

Taken together, our study provides an indication that manner adverbs influence motor behavior while corroborating the already existing data concerning the interaction between action verb processing and motor output. These findings are in line with assumptions made by embodied cognition theories proposing an essential role of sensorimotor areas in the processing and storage of action concepts inherent in action-related language. The adverbial modulation of motor behavior might reflect a certain variation of motor involvement in language processing. This involvement could be susceptible to grammatical constructions modifying the action component of action verbs. Yet, effects of the verb material in a closely matched verb set and influences of timing have to be taken into account.

**Acknowledgements** This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft). We thank Fabian Friedrich for the initial preparation of adverb stimuli, Anja Goldschmidt (SFB991/B09) and Thomas Gamerschlag (SFB991/Z) for linguistic advice, Tim Seuchter (SFB991-I-03/-II-B03) and Matthias Sure (SFB991/B03) for collecting parts of the data and Peter Indefrey (SFB991/A04, C03) for sharing expertise about priming experiments.

## References

- Andres, M., Finocchiaro, C., Buiatti, M., & Piazza, M. (2015). Contribution of motor representations to action verb processing. *Cognition*, 134, 174–184. <https://doi.org/10.1016/j.cognition.2014.10.004>.
- Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacobini, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology*, 16, 1818–1823. <https://doi.org/10.1016/j.cub.2006.07.060>.
- 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. <https://doi.org/10.1016/j.jml.2007.12.005>.
- Bak, T. H., O'Donovan, D. G., Xuereb, J. H., Boniface, S., & Hodges, J. R. (2001). Selective impairment of verb processing associated with pathological changes in Brodmann areas 44 and 45 in the motor neurone disease-dementia-aphasia syndrome. *Brain*, 124, 103–120. <https://doi.org/10.1093/brain/124.1.103>.

- Barsalou, L. W. (2008). Grounded cognition. *Annual Review Psychology*, 59, 617–645. <https://doi.org/10.1146/annurev.psych.59.103006.093639>.
- Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Biemann, C., Heyer, G., Quasthoff, U., & Richter, M. (2007). The Leipzig Corpora Collection—Monolingual corpora of standard size. In *Proceedings of Corpus Linguistics 2007*. Birmingham, UK.
- Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, 15(11), 527–536. <https://doi.org/10.1016/j.tics.2011.10.001>.
- Boulenger, V., Hauk, O., & Pulvermüller, F. (2009). Grasping ideas with the motor system: Semantic somatotopy in idiom comprehension. *Cerebral Cortex*, 19, 1905–1914. <https://doi.org/10.1093/cercor/bhn217>.
- Boulenger, V., Roy, A. C., Paulignan, Y., Deprez, V., Jeannerod, M., & Nazir, T. A. (2006). Cross-talk between language process and overt motor behavior in the first 200 msec of processing. *Journal of Cognitive Neuroscience*, 18, 1607–1615. <https://doi.org/10.1162/jocn.2006.18.10.1607>.
- Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research*, 24, 355–363. <https://doi.org/10.1016/j.cogbrainres.2005.02.020>.
- Dalla Volta, R., Gianelli, C., Campione, G. C., & Gentilucci, M. (2009). Action word understanding and overt motor behavior. *Experimental Brain Research*, 196(3), 403–412. <https://doi.org/10.1007/s00221-009-1864-8>.
- de Vega, M., Léon, I., Hernández, J. A., Valdés, M., Padrón, I., & Ferstl, E. C. (2014). Action sentences activate sensory motor regions in the brain independently of their status of reality. *Journal of Cognitive Neuroscience*, 26(7), 1363–1376. [https://doi.org/10.1162/jocn\\_a\\_00559](https://doi.org/10.1162/jocn_a_00559).
- Ehrenstein, W. H., & Arnold-Schulz-Gahmen, B. E. (1997). *Auge, Ohr, Hand und Fuß: Bestimmung des individuellen Lateralitätsprofils*. Dortmund: Institut für Arbeitsphysiologie.
- Fernandino, L., Conant, L. L., Binder, J. R., Blindsight, K., Hiner, B., Spangler, K., et al. (2013). Parkinson's disease disrupts both automatic and controlled processing of action verbs. *Brain and Language*, 127, 65–74. <https://doi.org/10.1016/j.bandl.2012.07.008>.
- Gianelli, C., & Dalla Volta, R. (2015). Does listening to action-related sentences modulate the activity of the motor system? Replication of a combined TMS and behavioral study. *Frontiers in Psychology*, 5(1511). <https://doi.org/10.3389/fpsyg.2014.01511>.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558–565. <https://doi.org/10.3758/BF03196313>.
- Goldschmidt, A., Gamerschlag, T., Petersen, W., Gabrovska, E., & Geuder, W. (2017). Towards verb modification in frames. In H. Hansen, S. Murray, M. Sadzadeh, & H. Zeevat (Eds.), *Logic, language, and computation. TbiLLC 2015. Lecture Notes in Computer Science* (Vol. 10148). Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-662-54332-0-2>.
- Grisoni, L., Dreyer, F. R., & Pulvermüller, F. (2016). Somatotopic semantic priming and prediction in the motor system. *Cerebral Cortex*, 26, 2353–2366. <https://doi.org/10.1093/cercor/bhw026>.
- Grossman, M., Anderson, C., Khan, A., Avants, B., Elman, L., & McCluskey, L. (2008). Impaired action knowledge in amyotrophic lateral sclerosis. *Neurology*, 71, 1396–1401. <https://doi.org/10.1212/01.wnl.0000319701.50168.8c>.
- Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41, 301–307. [https://doi.org/10.1016/S0896-6273\(03\)00838-9](https://doi.org/10.1016/S0896-6273(03)00838-9).
- Hauk, O., & Pulvermüller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping*, 21, 191–201. <https://doi.org/10.1002/hbm.10157>.
- Herrera, E., Rodríguez-Ferreiro, J., & Cuetos, F. (2011). The effect of motion content in action naming by Parkinson's disease patients. *Cortex*, 48(7), 900–904. <https://doi.org/10.1016/j.cortex.2010.12.007>.

- Kemmerer, D., Castillo, J. G., Talavage, T., Patterson, S., & Wiley, C. (2008). Neuroanatomical distribution of five semantic components of verbs: Evidence from fMRI. *Brain and Language*, 107, 16–43. <https://doi.org/10.1016/j.bandl.2007.09.003>.
- Kiefer, M., & Pulvermüller, F. (2011). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, 48(7), 805–825. <https://doi.org/10.1016/j.cortex.2011.04.006>.
- Klepp, A., Niccolai, V., Buccino, G., Schnitzler, A., & Biermann-Ruben, K. (2015). Language-motor interference reflected in MEG beta oscillations. *Neuroimage*, 109, 438–448. <https://doi.org/10.1016/j.neuroimage.2014.12.077>.
- Klepp, A., Niccolai, V., Sieksmeyer, J., Arnzen, S., Indefrey, P., Schnitzler, A., et al. (2017). Body-part specific interactions of action verb processing with motor behaviour. *Behavioural Brain Research*, 328, 149–158. <https://doi.org/10.1016/j.bbr.2017.04.002>.
- Klepp, A., Weissler, H., Niccolai, V., Terhalle, A., Geisler, H., Schnitzler, A., et al. (2014). Neuro-magnetic hand and foot motor sources recruited during action verb processing. *Brain and Language*, 128, 41–52. <https://doi.org/10.1016/j.bandl.2013.12.001>.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161–163. <https://doi.org/10.1038/307161a0>.
- Lange, J., Oostenveld, R., & Fries, P. (2011). Perception of the touch-induced visual doubleflash illusion correlates with changes of rhythmic neuronal activity in human visual and somatosensory areas. *Neuroimage*, 54, 1395–1405. <https://doi.org/10.1016/j.neuroimage.2010.09.031>.
- Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, 69(1), 1–33. <https://doi.org/10.18637/jss.v069.i01>.
- Maienborn, C., Alex-Ruf, S., Eikmeier, V., & Ulrich, R. (2015). Do we map remembrances to the left/back and expectations to the right/front of a mental timeline? Space-time congruity effects with retrospective and prospective verbs. *Acta Psychologica*, 156, 168–178. <https://doi.org/10.1016/j.actpsy.2014.11.006>.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164, 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>.
- Meteyard, L., Rodriguez Cuadrado, S., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, 48, 788–804. <https://doi.org/10.1016/j.cortex.2010.11.002>.
- Mollo, G., Pulvermüller, F., & Hauk, O. (2016). Movement priming of EEG/MEG brain responses for action-words characterizes the link between language and action. *Cortex*, 74, 262–276. <https://doi.org/10.1016/j.cortex.2015.10.021>.
- Moody, C. L., & Gennari, S. P. (2010). Effects of implied physical effort in sensory-motor and pre-frontal cortex during language comprehension. *NeuroImage*, 49, 782–793. <https://doi.org/10.1016/j.neuroimage.2009.07.065>.
- Moreno, I., de Vega, M., Léon, I., Bastiaansen, M., Lewis, A. G., & Magyari, L. (2015). Brain dynamics in the comprehension of action-related language. A time-frequency analysis of mu rhythms. *Neuroimage*, 109, 50–62. <https://doi.org/10.1016/j.neuroimage.2015.01.018>.
- Nakayashiki, K., Saeki, M., Takata, Y., Hayashi, Y., & Kondo, T. (2014). Modulation of event-related desynchronization during kinematic and kinetic hand movements. *Journal of Neuroengineering and Rehabilitation*, 11(1), 90. <https://doi.org/10.1186/1743-0003-11-90>.
- Niccolai, V., Klepp, A., Weissler, H., Hoogenboom, N., Schnitzler, A., & Biermann-Ruben, K. (2014). Grasping hand verbs: Oscillatory beta and alpha correlates of action-word processing. *PLoS ONE*, 9(9), e108059. <https://doi.org/10.1371/journal.pone.0108059>.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4).
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). Fieldtrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011. <https://doi.org/10.1155/2011/156869>.

- Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B*, 369(1651), 20130300. <https://doi.org/10.1098/rstb.2013.0300>.
- Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, 110, 1842–1857. [https://doi.org/10.1016/S1388-2457\(99\)00141-8](https://doi.org/10.1016/S1388-2457(99)00141-8).
- Pfurtscheller, G., Neuper, C., Andrew, C., & Edlinger, A. G. (1997). Foot and hand area mu rhythms. *International Journal of Psychophysiology*, 26, 121–135. [https://doi.org/10.1016/S0167-8760\(97\)00760-5](https://doi.org/10.1016/S0167-8760(97)00760-5).
- Postle, N., McMahon, K. L., Ashton, R., Meredith, M., & de Zubicaray, G. I. (2008). Action word meaning representations in cytoarchitectonically defined primary and premotor cortices. *Neuroimage*, 43, 634–644. <https://doi.org/10.1016/j.neuroimage.2008.08.006>.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6, 576–582. <https://doi.org/10.1038/nrn1706>.
- Pulvermüller, F. (2018). Neural reuse of action perception circuits for language, concepts and communication. *Progress in Neurobiology*, 160, 1–44. <https://doi.org/10.1016/j.pneurobio.2017.07.001>.
- Pulvermüller, F., Härtle, M., & Hummel, F. (2000). Neurophysiological distinction of verb categories. *NeuroReport*, 11(12), 2789–2793.
- Pulvermüller, F., Härtle, M., & Hummel, F. (2001). Walking or talking? Behavioral and neurophysiological correlates of action verb processing. *Brain & Language*, 78, 143–168. <https://doi.org/10.1006/brln.2000.2390>.
- Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neuroscience*, 21(5), 188–194. [https://doi.org/10.1016/S0166-2236\(98\)01260-0](https://doi.org/10.1016/S0166-2236(98)01260-0).
- Rüschemeyer, S. A., Lindemann, O., van Rooij, D., van Dam, W., & Bekkering, H. (2010). Effects of intentional motor actions on embodied language processing. *Experimental Psychology*, 57(4), 260–266. <https://doi.org/10.1027/1618-3169/a000031>.
- Sato, M., Mengarelli, M., Riggio, L., Gallese, V., & Buccino, G. (2008). Task related modulation of the motor system during language processing. *Brain and Language*, 105, 83–90. <https://doi.org/10.1016/j.bandl.2007.10.001>.
- Schacter, D. L., Wig, G. S., & Stevens, W. D. (2007). Reductions in cortical activity during priming. *Current Opinion in Neurobiology*, 17(2), 171–176. <https://doi.org/10.1016/j.conb.2007.02.001>.
- Scorolli, C., & Borghi, A. M. (2007). Sentence comprehension and action: Effector specific modulation of the motor system. *Brain Research*, 1130, 119–124. <https://doi.org/10.1016/j.brainres.2006.10.033>.
- Steingrüber, H. J. (2011). *Hand-Dominanz-Test*. Göttingen: Hogrefe.
- Tan, H., Pogosyan, A., Anzak, A., Ashkan, K., Bogdanovic, M., Green, A. L., et al. (2013). Complementary roles of different oscillatory activities in the subthalamic nucleus in coding motor effort in Parkinsonism. *Experimental Neurology*, 248, 187–195. <https://doi.org/10.1016/j.expneurol.2013.06.010>.
- Taylor, L. J., & Zwaan, R. A. (2008). Motor resonance and linguistic focus. *The Quarterly Journal of Experimental Psychology*, 61(6), 896–904. <https://doi.org/10.1080/17470210701625519>.
- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., et al. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, 17(2), 273–281.
- Ulrich, R., & Maienborn, C. (2010). Left-right coding of past and future in language. The mental timeline during sentence processing. *Cognition*, 117, 126–138. <https://doi.org/10.1016/j.cognition.2010.08.001>.
- van Dam, W. O., Rüschemeyer, S. A., & Bekkering, H. (2010). How specifically are action verbs represented in the neural motor system: An fMRI study. *Neuroimage*, 53, 1318–1325. <https://doi.org/10.1016/j.neuroimage.2010.06.071>.

- van Dijk, H., Nieuwenhuis, I. L., & Jensen, O. (2010). Left temporal alpha band activity increases during working memory retention of pitches. *European Journal of Neuroscience*, 31, 1701–1707. <https://doi.org/10.1111/j.1460-9568.2010.07227.x>.
- van Elk, M., van Schie, H. T., Zwaan, R. A., & Bekkering, H. (2010). The functional role of motor activation in language processing: Motor cortical oscillations support lexical-semantic retrieval. *Neuroimage*, 50, 665–677. <https://doi.org/10.1016/j.neuroimage.2009.12.123>.
- Yang, J., & Shu, H. (2011). Embodied representation of tool-use action verbs and hand action verbs: Evidence from a tone judgement task. *Neuroscience Letters*, 493, 112–115. <https://doi.org/10.1016/j.neulet.2011.02.020>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# When Mechanical Computations Explain Better



Silvano Zipoli Caiani

**Abstract** In this paper I defend the epistemic value of the representational-computational view of cognition by arguing that it has explanatory merits that cannot be ignored. To this end, I focus on the virtue of a computational explanation of optic ataxia, a disorder characterized by difficulties in executing visually-guided reaching tasks, although ataxic patients do not exhibit any specific disease of the muscular apparatus. I argue that addressing cases of patients who are suffering from optic ataxia by invoking a causal role for internal representations is more effective than merely relying on correlations between bodily and environmental variables. This argument has consequences for the epistemic assessment of radical enactivism, which invokes the Dynamical System Theory as the best tool for explaining cognitive phenomena.

**Keywords** Computational explanation · Dynamical system theory · Radical enactivism · Visual affordances · Optic ataxia

## 1 Introduction

According to a new generation of scholars, the *computational* paradigm that have informed the study of cognition for decades now creak under the weight of the new *enactivist* approach to cognition. Over the last few years, indeed, several philosophers and cognitive scientists have proposed to replace the *mechanical* and *representational* assumptions underlying the computational paradigm with a *dynamical* and *extensional* way to understand cognition (e.g., Chemero, 2011; Hutto & Myin, 2017; Gallagher, 2017). Supporters of the *radical enactivist view* (RE) argue that the computational paradigm does not add explanatory power over and above the physical

---

S. Zipoli Caiani (✉)  
Università Degli Studi Di Firenze, Florence, Italy  
e-mail: [silvano.zipolicaiani@unifi.it](mailto:silvano.zipolicaiani@unifi.it)

description of a cognitive system, and therefore it should be abandoned (e.g., van Gelder, 1995; Chemero, 2011; Hutto & Myin, 2012).<sup>1</sup>

The aim of this paper is to defend the epistemic value of the computational view of cognition by arguing that it has explanatory merits that cannot be ignored. To this end, I focus on the virtue of a mechanical-computational explanation of the behavior of patients suffering from *optic ataxia*, a disorder characterized by difficulties in executing visually-guided reaching tasks, although patients do not exhibit any specific disease of the muscular apparatus (Balint, 1909). I argue that addressing cases of patients who are suffering from optic ataxia by invoking the causal role for internal representations is more effective than merely relying on correlations between bodily and environmental variables.

According to the computational paradigm, the cognitive system forms visual representations of the available actionable opportunities in the environment, which have a causal role in action planning and execution (e.g., Fodor & Pylyshyn, 1988; McCulloch & Pitts, 1944; Putnam, 1967). This serves to emphasize the need to identify the parts and the mechanical structures characterizing the causal chains underlying and generating the behavior of interest (Craver & Darden, 2013; Illari & Williamson, 2012; Bechtel & Richardson, 1993; Craver, 2006). Thus, such an account shows why an agent performs a certain behavior by describing the relevant mechanisms linking internal representations with the agent's motor system.<sup>2</sup>

In a different vein, RE denies the need to invoke internal representations to account for the interaction between vision and action. According to RE, modeling the relationships between vision and action requires attending to the ways in which individuals dynamically engage with certain worldly offerings by means of extended interactions (Hutto & Myin, 2017). In doing this, RE assumes that visual cognition does not involve the selecting, storing, and processing of information in the brain. Differently, RE conceives visual cognition as an *extensive phenomenon* concerning the variation of bodily and environmental variables spanning multiple temporal and spatial scales. This amounts to an assumption that the agent and the environment form a unified system whose behavior cannot be modeled as a causal chain linking separate parts (Chemero, 2011). Accordingly, the interlocking between vision and action should be explained via a methodological framework that does not posit mental representations, like *dynamical systems theory* (DST). Notably, modeling cognition by means of DST allows for a lawful account of how agents interact with the action-related properties of the environment, without the need to involve internal resources such as causal states and computations (e.g., Beer, 2000; Spivey, 2008; Chemero, 2011).

---

<sup>1</sup>For the sake of the present argument, I focus exclusively on Radical Enactivism (e.g., Chemero, 2011; Hutto & Myin, 2012), excluding the different theoretical strands that populate the enactivist world (e.g., Maturana & Varela, 1991; Noë, 2004; O'Regan & Noë, 2001; O'Regan, 2011). At present, radical enactivism is the most developed, discussed, and challenging alternative to the classical computational paradigm that has informed cognitive science for about sixty years.

<sup>2</sup>It should be noted that the mechanical approach to explanation improves our comprehension of the causal chain that allows a behavior to occur in conjunction with certain environmental conditions, thus making the execution of an action a *non-surprising* event (Cohen, 2015; Schupbach & Sprenger, 2011).

Though RE is currently encountering enthusiastic appraisals, it is not un-common that someone may still consider it as a proposal that is more easily explained than proved. Whether RE is only on the crest of a fashionable wave that is doomed to leave no tracks in the sand, or whether it is a tsunami with the power to sweep away the existing explanatory practice in the cognitive sciences, is something that has not yet been carefully assessed. In order to address this issue, I follow Hutto and Myin in considering that the only naturalistically respectable way to decline RE is “to give it its day in empirical court” (Hutto & Myin, 2017, p. 19). This amounts to wondering whether the methodological tools of DST, instead of a computational-mechanical approach, offer the best explanation of basic cognitive phenomena. This paper shows that there are factual circumstances concerning the ability and inability to perceive and exploit visual affordances for which DST is not able to explain, but which are suitably accounted for through the adoption of a computational architecture based on the dual streams model of vision (Goodale & Milner, 1992). More precisely, I maintain that RE provides a valuable explanation of *why* agents perceive action opportunities, whereas a computational view provides an explanation for *why* agents have such an ability in addition to *why* this ability can be lost.

This paper is divided into six parts. In the first part (Sect. 2), I introduce RE by distinguishing between two claims, the former concerning the ontological status of representational entities, and the latter concerning the explanatory power of a non-representational account of cognition. In Sect. 3, I focus on the explanatory claim and provide details concerning the strategy underlying DST, showing that it amounts to a correlational approach. In Sect. 4, I introduce the case of optic ataxia, and argue that it is an ideal target for measuring the explanatory power of DST. In Sect. 5, I show that the correlational analysis provided by DST is not suitable to explain relevant aspects of ataxic behavior, since it does not suffice to provide an etiological account for it. Finally, in Sect. 6, I introduce a computational model of vision for action, and show that it is suitable to provide the etiological account that is required in the case of optic ataxia.

As a result, although sometimes the dynamical systems theory and the computational paradigm can be “natural allies”, playing both a complementary role in describing the interactions between vision and action (Kaplan, 2015), in the case of optic ataxia, the computational view is more explanatory than the dynamical one. This outlines that there is an epistemological shortcoming of radical enactivism compared to the computational account.

## 2 Radical Enactivism and the Explanatory Claim

According to RE, there are cognitive facts that can be fully and completely accounted for by means of an *extensional* language, that is, by conceiving them merely in terms of activities in which the agent’s body is dynamically engaged with the environment. Notably, considering cognitive phenomena in a purely extensional way, supporters

of RE state that the body-environment relations do not involve any computational manipulation of information (e.g., Chemero, 2011; Hutto & Myin, 2012, 2017).

In denying the computational nature of cognition, supporters of RE might be committed to more than one claim. As Chemero (2011) has noted, when one proclaims that cognition does not involve computations, there are at least two theoretical views one might endorse. First, one might be making a claim about *what there is* and *what there is not*, namely, a claim about the ontology of the cognitive sciences. Second, one might be claiming something about the best way to provide explanatory arguments in the cognitive sciences. While in the former case, the rebuttal of the computational view amounts to a *metaphysical* thesis, in the latter case it rests on *epistemological* grounds, that is, on the analysis of the needs and practices that characterize the work of cognitive scientists. The key difference between the two claims is that only the explanatory claim is an empirical hypothesis, whereas the metaphysical claim concerns our philosophical criteria for establishing the place of cognition in nature (Chemero, 2011).

Over the last decades, many arguments have been raised against the attempt to provide a successful naturalization of computational systems (for a review see, Kriegel, 2013; Pietroski, 1992; Ramsey, 2007, 2015), such that it is an ongoing debate whether computational processes should be considered parts of the natural ontology or not. Although it raises a fascinating philosophical discussion, the metaphysical hypothesis has little impact on the scientific practice since one may continue to refer to a computational approach to cognition with or without compromising with any sort of naturalization of the computational states (classically Dennett, 1987; more recently see Egan, 2013; Colombo, 2014). Accordingly, given the different purposes underlying the practical use of a word such as “computations”, the metaphysical claim is hardly defensible on empirical grounds (Chemero, 2011).

Differently, the epistemological hypothesis concerning the explanatory value of the computational approach to cognition has dramatic consequences on the real practices of cognitive scientists. According to this hypothesis, the great variety of experiences and behaviors are best understood without appealing to the manipulation of causal states but rather by focusing on the dynamical interactions between the agent's body and the environment. When it comes to accounting for intelligent activity, supporters of RE subscribe to the *Equal Partner Principle* (Hutto & Myin, 2017), according to which variables of any kind make an equal explanatory contribution, regardless of whether they concern aspects located in or out of the boundaries of skull and skin. This means that citing internal factors endowed with a causal status does not carry more explanatory value than, for example, referring to environmental and bodily factors that merely correlate with each other. Accordingly, since the computational view is refuted as an explanatory tool for the cognitive sciences, agents and environmental factors can be modeled as a unified, non-decomposable system whose behavior cannot be accounted for, even approximately, as a set of separate causal parts.

### 3 Radical Enactivism and the Dynamical System Theory

According to the previous considerations, the adoption of DST may be pivotal for an epistemological approach to RE (e.g., Beer, 2000; Chemero, 2011; van Gelder, 1995; Heinke, 2000; Walmsley, 2008). Indeed, the methodological assumptions underlying DST allow for an approach to the study of cognition that avoids mechanical states and inner computational processing (Spivey, 2008, Chemero, 2011). Conceiving cognition from an extensional point of view allows for a lawful account of how agents interact with the action-related properties of the environment, without the need to involve internal resources such as causal states and computations. According to DST, cognitive explanations are arguments based on factual premises and inferential rules inasmuch as they take the form of a reasoning in which the phenomenon to explain (*explanandum*) follows as a deductive consequence of the selected premises (*explanans*). This is, indeed, the core idea of the well-known *covering-law* model of explanation (Hempel, 1965; Walmsley, 2008).<sup>3</sup>

Over the last few decades, this methodological approach has been endorsed in the cognitive science of vision to account for the way agents perceive an *affordance* in the environment, that is, a *possibility of action* that surround the agent's body (Gibson, 1979). According to this view, the perception of affordances is construed as the detection of a *relation* between features of the environment and certain motor-related properties of the agent's body. Hence, in order to study the perception of affordances by means of DST, some environmental parameters should be considered in relation to some relevant variables concerning the agent's body and the related motor skills (e.g., Harrison, Turvey, & Frank, 2016; Lopresti-Goodman, Turvey, & Frank, 2011; Mark, 1987; Rietveld & Kiverstein, 2014).

Therefore, if a cognitive agent guides its activity by detecting affordances in the environment, it is possible to suppose that these affordances must be sensible with regard to the *lawful relationships* between environmental aspects and the relevant features and motor skills of its own body. DST, indeed, starts by selecting the critical parameters that characterize the state of the agent-environment system and attempts to disclose the way such parameters relate with one another. Then, DST focuses on the trajectories in a phase space that the parameters of the agent-environment system traverses, given the covariation of bodily, practical and environmental variables, describing the laws according to which its behavior changes because of the modification of one or more parameters (Beer, 2000; Chemero, 2011).

To this extent, DST improves our access to the laws governing the interactions between the agent's body and its environment, thus making the occurrence of a certain agent's intelligent behavior not a surprising event. This would be particularly evident if we were interested in making predictions concerning the manner in which agents'

---

<sup>3</sup>In this view, one explains the occurrence of a certain event E by arguing that it is expected because of the factual conditions  $C_1 \dots C_n$  and the deductive laws  $L_1 \dots L_n$ . Such a type of explanation is suitable to answer the question "Why does phenomenon E occur?" by showing that its occurrence—or its probability of occurring—results from the combination of particular circumstances ( $C_1 \dots C_n$ ), in accordance with the general laws ( $L_1 \dots L_n$ ).

behavior varies over time. Indeed, once we know the relevant ambient parameters and the laws governing the dynamical evolution of the environment-agent system, the future values of the agent's behavior become nothing but a matter of deduction. According to this view, if a dynamical systems account is sufficiently accurate to describe what would occur in counterfactual circumstances, it can be considered as a tool suitable for reducing surprise about the occurrence of a behavioral event (e.g., Chemero & Silberstein, 2008; Thelen & Smith, 1996).

To sum up, the epistemological approach to RE and the methodological tools of DST form a joint venture that has recently attracted the attention of an increasing number of cognitive scientists. DST, indeed, rests on the *Equal Partner Principle* by providing an account of the agent's behavior that does not discriminate between internal and external resources. Accordingly, DST offers deductive-nomological explanations that are merely based on the fine-grained analysis of the internal dynamics characterizing the covariation of selected parameters spanning the agent's brain, the body and its environment.

Although RE is gaining an increasing consensus, it is still an open issue whether it will be able to replace the mechanical-computational paradigm that has guided the cognitive sciences over the last sixty years. If so, DST should be able to provide a satisfactory explanation of any sort of cognitive phenomena, with emphasis on the agent's *basic cognitive behaviors*, such as the perception and misperception of affordances in the environment (Hutto & Myin, 2012). However, in the remaining part of this paper, I will show that this is not the case.

## 4 Explaining Anomalies: The Case of Optic Ataxia

The study of cognition is not a mere theoretical game, but it has relevant practical implications for the development of therapies and rehabilitation programs for patients suffering from cognitive deficits. Considering this purpose, it is interesting to assess the explanatory virtue of RE as it pertains to its possible clinical consequences. Thus, it may be helpful to assess the adoption of DST as a methodological tool for the explanation of non-standard cases of perception such as *optic ataxia*, a condition in which some or all aspects of visual guidance of reaching with the hand and arm are lost. Patients suffering from optic ataxia have an intact visual field, good oculomotor control, and normal motor skills; however, they are not able to detect the possible practical relations between their motor abilities and the features of the environment, meaning that they are not able to perceive the affordances available to them.

The scientific literature concerning cases of optic ataxia reports alterations in the initial and final stages of the visually guided movement of reaching to grasp. Anomalous dynamics have been reported in scaling the aperture of the hand according to the target (Cavina-Pratesi, Connolly, & Milner, 2013) in following objects' trajectories and in executing the final stage of a grasping action (Blangero et al., 2010). Furthermore, ataxic patients show a lack of automatic correction when a target changes location (Pisella et al., 2000) and a lack of ability to avoid collisions with distractors

when reaching for something (Schindler et al., 2004). However, although optic ataxia is a permanent impairment, patients can relieve their deficit and improve their performances by means of specific rehabilitation programs. For example, patients exhibit an enhanced performance in reaching and grasping when a delay is introduced between the perceptual stimulus and the behavioral response (Himmelbach & Karnath, 2005). Moreover, a common rehabilitation program includes compensatory strategies such as the recourse to external prostheses (e.g., planners, calendars, recording devices, timers and pagers) in addition to internal cueing (e.g., developing mnemonics or an internal checklist). Generally, patients have been demonstrated to reduce errors and improve performance by following non-perceptual cues, such as conceptual information, but only when their memory is relatively preserved (Zgaljardic et al., 2011).

Evidence such as this begs for an explanation. Notably, two main questions arise: the first concerns the very etiology of optic ataxia, and the second concerns the fact that, at least in certain cases, ataxic patients exhibit good performance. It is interesting, indeed, to understand *why* patients with lesions precisely located in the parietal cortex are not able to detect and select affordances in the environment and *why* precisely the execution of delayed tasks and the retrieval of conceptual information improve patient performances (Himmelbach & Karnath, 2005; Zgaljardic et al., 2011). To this extent, explaining optic ataxia may be used as a testing ground for examining the epistemic virtue of DST. It seems reasonable, indeed, to assume that a good account of basic cognitive abilities should be able to address anomalous cases as well. Accordingly, a valuable explanation of affordance perception should explain *why* agents may lose such an ability as well as *why* they may be able to recover it given certain circumstances.

## 5 Covariation Is not Enough

Because DST approaches the perception of the affordances by means of covering-law explanations (see Sect. 3), it provides an account of optic ataxia that is addressed on the covariation of selected parameters that characterize the state of the agent-environment system. Notably, in explaining the anomalous behavior of patients suffering from optic ataxia, DST focuses on the trajectories in a mathematical *phase space* that the agent-environment system traverses over time, and it specifies how they depend on changes in one or more parameters of the coupled system.

The efforts of scholars working in the context of DST has been merely devoted to observe how patterns of correlation between bodily and environmental variables emerge, stabilize and are sometimes lost. Indeed, according to a correlational approach, explaining anomalous performances in perceiving and exploiting affordances requires the identification of appropriate patterns of variables to quantify and qualify the nature of the deficit. Although DST is usually focused on non-disabled individuals, several studies have recently measured the ability to perform visually guided reaching actions in patients with lesions to the parietal cortex that are

comparable to those characteristics of optic ataxia (e.g., Kamper et al., 2002; Pisella, Rossetti, & Rode, 2017).<sup>4</sup>

Correlational evidence provides quantitative data to assess the actual disruption of default modes of coordination in ataxic patients and the possible motor-control gain following rehabilitation therapy. The available experiments show that ataxic patients do not detect the dynamical relationship between environmental features and the motor properties of their own bodies. This means that ataxic patients are unable to judge the scaling of environmental variables in relation to their bodily variables, resulting in the performance of anomalous behavioral patterns.

However, although the study of correlational variables provides a description of the dysfunctional ataxic behavior, this approach offers no cues concerning *the causes* underlying such conditions. This means that a correlational account can be fruitfully employed to gain information about the variability of the disease symptoms, showing different degrees of severity with respect to standard behavioral patterns, but it cannot be employed for the purpose of *etiological diagnosis*. Indeed, the methodological tools of DST are not suitable for highlighting the individual causes of a disease phenomenon (see Sect. 3), thus DST is unable to explain *why* ataxic patients with are impaired in performing visually guided grasping actions. After a complete correlational analysis, one may still require an explanation of *why lesions in the parietal cortex correlate with ataxic behaviors*, albeit no correlational analysis can answer this question. Though a correlational methodology allows one to predict that lesions in the parietal cortex usually result in the inability of the agent to detect action possibilities in the environment, it seems incompetent in explaining *why* there are cases in which they reduce errors and relieve their conditions.

Of course, a correlational approach may be able to predict this phenomenon by means of generalizations based on previous cases but is unable *to say why* such a phenomenon occurs. A correlational account, indeed, is unable *to explain why* using conceptual information may improve the performance of ataxic patients (Zgaljardic et al., 2011). The mere knowledge that an alteration of cortical parameters is correlated with variations in parameters concerning visually guided actions does not provide sufficient reasons to infer that the recourse to external prostheses (e.g., planners, calendars, recording devices, timers and pagers) in addition to internal cueing (e.g., developing mnemonics or an internal checklist) may reduce errors and relieve

---

<sup>4</sup>Experimental results show that after a measurable lesion in the left posterior parietal cortex, the agent's ability to reach a target is characterized by significant alterations in several parameters such as the initial movement direction, decreased hand velocity, decreased elbow velocity, and increased trajectory curvature (Kamper et al., 2002). A purely correlational analysis also shows that patients with lesions to the parietal cortex have difficulty in performing reaching-to-grasp actions located in the contralesional visual field and with the contralesional hand. In this respect, a relevant discrepancy is observed when ataxic patients use the ataxic hand for actions directed towards the ataxic field, whereas less severe discrepancies are observed when patients use the healthy hand towards the ataxic visual field or the ataxic hand towards the healthy visual field. In contrast, actions performed with the healthy hand towards the healthy visual field exhibit no discrepancies compared with normal subjects (Pisella, Rossetti, & Rode, 2017).

the condition of patients suffering from optic ataxia (Sect. 4). This means that a rehabilitation program based on such a kind of resource is hardly configurable from the point of view of DST, and its results cannot be explained by means of a correlational approach.

## 6 When Computations Explain Better

Over the last few decades, the dual stream model of visual processing (Jacob & Jeannerod, 2003; Milner & Goodale, 1995) served as basic ground to build up a computational architecture according to which an agent computes visuomotor information in the environment (e.g., Cisek & Kalaska, 2010; Thill et al., 2013; Zipoli Caiani, 2014; Tillas et al., 2017). According to the dual streams model, visual processing involves two subsystems: the *dorsal system*, which performs processes associated with detecting affordances and visually guiding actions, and the *ventral system*, which performs processes associated with semantic identification and intentional planning (Goodale & Milner, 1992).

The essence of the dual streams model of vision lies in the functional differences between the two streams. On one hand, the ventral stream allows an agent to recognize objects in the environment, attaching meanings and establishing causal relations. Such operations are crucial for acquiring a *conceptual grasp* of the environment, providing resources for incorporating previously stored information into the online control of current actions and making intentional action planning possible (Goodale & Milner, 1995; Goodale, 2014). On the other hand, the dorsal stream performs transformations that convert information about the shape and location of the source of the stimulus into parameters suitable for action execution. Along the dorsal pathway, the anterior intraparietal area and the ventral premotor cortex extract and compute sensorimotor information from the perceptual stimulus, making it possible to detect action possibilities from the information detected through the retinotopic map (e.g., Andersen & Buneo, 2003; Mohan et al., 2017; Rizzolatti & Luppino, 2001).<sup>5</sup>

Importantly, over the last few years, several studies have shown that the ventral stream also biases the *detection* of action possibilities by exploiting *functional interactions* with different points of the dorsal processing (Briscoe, 2009; Briscoe & Schwenkler, 2015; Chinellato & Pobil, 2016; Zipoli Caiani & Ferretti, 2017). Among the various interactions between the information processed in the two streams, an

---

<sup>5</sup>A generally agreed-upon architecture for affordance perception assumes that visuomotor information is computed by means of a *sensorimotor matching mechanism* (Rizzolatti & Sinigaglia, 2008). This amounts to an assumption that action-related information is detected and processed by the agent's sensorimotor apparatus depending on its body shape and motor abilities. According to this view, since the stimulus information in visual perception and the motor information underlying the action are coded together (Prinz, 1997), it seems possible to account for the attentional facilitation that characterizes the detection of action possibilities in terms of visually elicited *motor representations* (Brozzo, 2017; Butterfill & Sinigaglia, 2014; Ferretti, 2016, Ferretti & Zipoli Caiani, 2019).

important connection is precisely that which occurs at the level of the parietal cortex, that is, the region of the brain damaged in optic-ataxic patients. This interaction strongly affects motor preparation and control of movements, suppressing elicited sensorimotor patterns to prevent undesired actions from being triggered. Indeed, information from the ventral stream may help in selecting the relevant patterns of action processed in the dorsal pathway, allowing to the agent's conceptual knowledge to influence the execution of visually guided actions (e.g., Borra et al., 2010; Hoshi & Tanji, 2007). It may be argued that this computational architecture offers an adaptive advantage to the extent that it allows a fast link between perception, conceptualization and action by means of reliable information integration (Zipoli Caiani, 2018).

Concerning the computational role of the parietal cortex, emerging data from neuropsychology and neuroimaging support the view that portions of this region are devoted to integrating information for guiding actions according to the agent's specific goal (Culham, Cavina-Pratesi, & Singhal, 2006). Notably, a number of TMS studies have shown that the parietal cortex is functionally involved in the processing of the visual motor information required to adjust the motor plan to perform hand actions and achieve intentional goals (Iacoboni, 2006). Evidence such as this shows that the parietal cortex is responsible for representation and conversion of visual information into movements and for online control of motor actions (Blangero et al., 2010). Lesions in this area, therefore, leave patients without a fundamental structure for visuomotor integration and control, thus causing disorders in the representation of the surrounding objects and impairments in planning and execution of goal-directed actions.

Interestingly, the computational architecture based on the dual stream model of vision *explains why* patients with lesions to the parietal cortex may suffer from optic ataxia. Moreover, the same architecture *explains why* ataxic patients exhibit intact performance when a delay is introduced between the perceptual stimulus and the behavioral response or when the patient relies on conceptual knowledge of the target. According to this architecture, the impairment of the parietal cortex does not completely prevent agents from processing and exploiting conceptual information. The massive interaction between the ventral stream and the dorsal stream allows for a *reallocation of functions* that ensures the recognition of affordances in the environment by means of compensatory strategies such as the exploitation of conceptual cues (Zipoli Caiani & Ferretti, 2017). This means that, once the functional specializations and reciprocal interactions between the two streams have been defined, it is possible to *explain why* a lesion in the parietal cortex may induce the inability to *immediately detect* a pragmatic relation between the agent's body and the features of the environment. Moreover, it is also possible to *explain why*, in certain circumstances, the use of conceptual information may relieve such a deficit.

## 7 Conclusions

RE is a view according to which cognitive phenomena should be explained by means of DST instead of a mechanical computational account. Although the adoption of the methodological tools of DST is gaining increasing consensus in the cognitive science of vision, it faces an explanatory shortcoming that should not be underestimated. It is well known, indeed, that *descriptive* and *predictive adequacy* do not imply *explanatory adequacy* (Salmon, 1984). Accordingly, although DST is suitable to provide precise correlations between environmental, bodily and behavioral variables, such a methodology remains silent about the underlying causes of such correlations.

However, by means of a computational architecture based on the dual stream model of vision it has been possible to explain *why* patients with lesions to the parietal cortex become unable to detect affordances in the environment, as well as *why* they gain good visuomotor performances given appropriate conditions (see Sect. 6). The computational integration of pragmatic and conceptual information in vision for action, indeed, makes it possible to explain *why* a lesion in an area of the parietal cortex is correlated with inability to detect and exploit affordances, but also *why* particular circumstances (e.g., delayed responses and conceptual information) allow the agent to use alternative cognitive strategies to recognize and take advantage of the affordances of the environment.

## References

- Andersen, R. A., & Buneo, C. A. (2003). Sensorimotor integration in posterior parietal cortex. *Advances in Neurology*, 93, 159–177.
- Balint, R. (1909). Seelenlähmung des. *Monatsschr Psychiatr Neurol*, 25, 51–81.
- Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity decomposition and localization as strategies in scientific research*. Princeton.
- Beer, (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, 4(3), 91–99.
- Blangero, A., Ota, H., Rossetti, Y., Fujii, T., Ohtake, H., Tabuchi, M., et al. (2010). Systematic retinotopic reaching error vectors in unilateral optic ataxia. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 46(1), 77–93. <https://doi.org/10.1016/j.cortex.2009.02.015>.
- Borra, E., Ichinohe, N., Sato, T., Tanifugi, M., & Rockland, K. S. (2010). Cortical connections to area TE in monkey: hybrid modular and distributed organization. *Cerebral Cortex (New York, N.Y.: 1991)*, 20(2), 257–270. <https://doi.org/10.1093/cercor/bhp096>.
- Briscoe, R. (2009). Egocentric spatial representation in action and perception. *Philosophy and Phenomenological Research*, 79(2), 423–460.
- Briscoe, R., & Schwenkler, J. (2015). Conscious vision in action. *Cognitive Science*, 39(7), 1435–1467.
- Brozzi, C. (2017). Motor intentions: how intentions and motor representations come together. *Mind Language*, 32(2), 231–256. <https://doi.org/10.1111/mila.12140>.
- Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in Human Neuroscience*, 8, 599. <https://doi.org/10.3389/fnhum.2014.00599>.

- Butterfill, S. A., & Sinigaglia, C. (2014). Intention and motor representation in purposive action: intention and motor representation in purposive action. *Philosophy and Phenomenological Research*, 88(1), 119–145. <https://doi.org/10.1111/j.1933-1592.2012.00604.x>.
- Cavina-Pratesi, C., Connolly, J. D., & Milner, A. D. (2013). Optic ataxia as a model to investigate the role of the posterior parietal cortex in visually guided action: evidence from studies of patient M.H. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00336>.
- Chemero, A. (2011). *Radical embodied cognitive science*. Bradford.
- Chemero, T., & Silberstein, M. (2008). After the philosophy of mind: replacing scholasticism with science. *Philosophy of Science*, 75(1), 1–27.
- Chinellato, E., & del Pobil, A. P. (2016). The neuroscience of action and perception. In *The visual neuroscience of robotic grasping* (pp. 7–38). Springer, Cham. [https://doi.org/10.1007/978-3-319-20303-4\\_2](https://doi.org/10.1007/978-3-319-20303-4_2).
- Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, 33, 269–298. <https://doi.org/10.1146/annurev.neuro.051508.135409>.
- Cohen, M. P. (2015). On Schupbach and Sprenger's measures of explanatory power. *Philosophy of Science*, 82(1), 97–109. <https://doi.org/10.1086/678980>.
- Colombo, M. (2014). Neural representationalism, the hard problem of content and vitiated verdicts. A reply to Hutto Myin. *Phenomenology and the Cognitive Sciences*, 13(2), 257–274.
- Craver, C. F. (2006). When mechanistic models explain. *Synthese*, 153(3), 355–376.
- Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. University of Chicago Press.
- Culham, J. C., Cavina-Pratesi, C., & Singhal, A. (2006). The role of parietal cortex in visuo-motor control: what have we learned from neuroimaging? *Neuropsychologia*, 44(13), 2668–2684. <https://doi.org/10.1016/j.neuropsychologia.2005.11.003>.
- Dennett, D. C. (1987). *The intentional stance*. MIT Press.
- Egan, F. (2013). How to think about mental content. *Philosophical Studies*, 1, 1–21.
- Ferretti, G. (2016). Through the forest of motor representations. *Consciousness and Cognition*, 43, 177–196. <https://doi.org/10.1016/j.concog.2016.05.013>.
- Ferretti, G., & Zipoli Caiani, S. (2019). Solving the interface problem without translation: the same format thesis. *Pacific Philosophical Quarterly*, 100(1), 301–333.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture. *Cognition*, 28(1–2), 3–71.
- Gallagher, S. (2017). *Enactivist interventions: rethinking the mind* (1st ed.). Oxford, United Kingdom: Oxford University Press.
- van Gelder, T. (1995). What might cognition be if not computation? *Journal of Philosophy*, 92(7), 345–381.
- Gibson, J. J. (1979). *The ecological approach to visual perception: Classic edition*. Psychology Press.
- Goodale, M. A. (2014). How (and why) the visual control of action differs from visual perception. *Proceedings. Biological Sciences*, 281(1785), 20140337. <https://doi.org/10.1098/rspb.2014.0337>.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20–25.
- Harrison, H. S., Turvey, M. T., & Frank, T. D. (2016). Affordance-based perception-action dynamics: A model of visually guided braking. *Psychological Review*, 123(3), 305–323. <https://doi.org/10.1037/rev0000029>.
- Heinke, D. (2000). A dynamical system theory approach to cognitive neuroscience. *Behavioral and Brain Sciences*, 23(4), 543.
- Hempel, C. G. (1965). *Aspects of scientific explanation: and other essays in the philosophy of science*. Free Press.

- Himmelbach, M., & Karnath, H.-O. (2005). Dorsal and ventral stream interaction: contributions from optic ataxia. *Journal of Cognitive Neuroscience*, 17(4), 632–640. <https://doi.org/10.1162/0898929053467514>.
- Hoshi, E., & Tanji, J. (2007). Distinctions between dorsal and ventral premotor areas: anatomical connectivity and functional properties. *Current Opinion in Neurobiology*, 17(2), 234–242. <https://doi.org/10.1016/j.conb.2007.02.003>.
- Hutto, D. D., Myin, E. (2012). *Radicalizing enactivism: Basic minds without content*. MIT Press.
- Hutto, D. D., Myin, E. (2017). *Evolving enactivism: Basic minds meet content*. MIT Press. Retrieved from <http://www.jstor.org/stable/j.ctt1q1xq5g>.
- Iacoboni, M. (2006). Visuo-motor integration and control in the human posterior parietal cortex: evidence from TMS and fMRI. *Neuropsychologia*, 44(13), 2691–2699. <https://doi.org/10.1016/j.neuropsychologia.2006.04.029>.
- Illari, P., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science*, 2(1), 119–135.
- Jacob, P., Jeannerod, M. (2003). *Ways of seeing: The scope and limits of visual cognition*. Oxford University Press.
- Kamper, D. G., McKenna-Cole, A. N., Kahn, L. E., & Reinkensmeyer, D. J. (2002). Alterations in reaching after stroke and their relation to movement direction and impairment severity. *Archives of Physical Medicine and Rehabilitation*, 83(5), 702–707.
- Kaplan, D. M. (2015). Moving parts: The natural alliance between dynamical and mechanistic modeling approaches. *Biology and Philosophy*, 30(6), 757–786.
- Kriegel, U. (2013). The phenomenal intentionality research program. In U. Kriegel (Ed.), *Phenomenal intentionality*. Oxford University Press.
- Lopresti-Goodman, S. M., Turvey, M. T., & Frank, T. D. (2011). Behavioral dynamics of the affordance “graspable”. *Attention, Perception, & Psychophysics*, 73(6), 1948–1965. <https://doi.org/10.3758/s13414-011-0151-5>.
- Mark, L. S. (1987). Eyeheight-scaled information about affordances: a study of sitting and stair climbing. *Journal of Experimental Psychology. Human Perception and Performance*, 13(3), 361–370.
- Maturana, H. R., & Varela, F. J. (1991). *Autopoiesis and Cognition: The realization of the living*. Springer Science & Business Media.
- Mcculloch, W. S., & Pitts, W. (1944). A logical calculus of the ideas immanent in nervous activity. *Journal of Symbolic Logic*, 9(2), 49–50.
- Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford University Press.
- Mohan, H., de Haan, R., Mansvelder, H. D., & de Kock, C. P. J. (2017). The posterior parietal cortex as integrative hub for whisker sensorimotor information. *Neuroscience*. <https://doi.org/10.1016/j.neuroscience.2017.06.020>.
- Noë, A. (2004). *Action in perception*. MIT Press.
- O'Regan, J. K. (2011). *Why red doesn't sound like a bell: Understanding the feel of consciousness*. Oxford University Press.
- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *The Behavioral and Brain Sciences*, 24(5), 939–973; discussion 973–1031.
- Paolo, E. D. (2009). Extended life. *Topoi*, 28(1), 9–21.
- Pietroski, P. M. (1992). Intentionality and teleological error. *Pacific Philosophical Quarterly*, 73(3), 267–282.
- Pisella, L., Gréa, H., Tilikete, C., Vighetto, A., Desmurget, M., Rode, G., et al. (2000). An “automatic pilot” for the hand in human posterior parietal cortex: toward reinterpreting optic ataxia. *Nature Neuroscience*, 3(7), 729–736. <https://doi.org/10.1038/76694>.
- Pisella, L., Rossetti, Y., & Rode, G. (2017). Optic ataxia in Bálint-Holmes syndrome. *Annals of Physical and Rehabilitation Medicine*, 60(3), 148–154. <https://doi.org/10.1016/j.rehab.2016.01.003>.
- Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, 9(2), 129–154. <https://doi.org/10.1080/713752551>.

- Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, mind, and religion* (pp. 37–48). University of Pittsburgh Press.
- Ramsey, W. (2015). Must cognition be representational? *Synthese*. <https://doi.org/10.1007/s11229-014-0644-6>.
- Ramsey, W. M. (2007). *Representation reconsidered*. Cambridge University Press.
- Rietveld, E., & Kiverstein, J. (2014). A rich landscape of affordances. *Ecological Psychology*, 26(4), 325–352. <https://doi.org/10.1080/10407413.2014.958035>.
- Rizzolatti, G., & Luppino, G. (2001). The cortical motor system. *Neuron*, 31(6), 889–901.
- Rizzolatti, G., & Sinigaglia, C. (2008). *Mirrors in the brain: How our minds share actions and emotions*. Oxford University Press.
- Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.
- Schindler, I., Rice, N. J., McIntosh, R. D., Rossetti, Y., Vighetto, A., & Milner, A. D. (2004). Automatic avoidance of obstacles is a dorsal stream function: evidence from optic ataxia. *Nature Neuroscience*, 7(7), 779–784. <https://doi.org/10.1038/nn1273>.
- Schupbach, J. N., & Sprenger, J. (2011). The Logic of Explanatory Power\*. *Philosophy of Science*, 78(1), 105–127. <https://doi.org/10.1086/658111>.
- Spivey, M. (2008). *The continuity of mind*. Oxford, New York: Oxford University Press.
- Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*. MIT Press.
- Thill, S., Caligiore, D., Borghi, A. M., Ziemke, T., & Baldassarre, G. (2013). Theories and computational models of affordance and mirror systems: an integrative review. *Neuroscience and Biobehavioral Reviews*, 37(3), 491–521. <https://doi.org/10.1016/j.neubiorev.2013.01.012>.
- Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.
- Tillas, A., Vosgerau, G., Seuchter, T., & Zipoli Caiani, S. (2017). Can affordances explain behavior? *Review of Philosophy and Psychology*, 8(2), 295–315.
- Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT Press.
- Walmsley, J. (2008). Explanation in dynamical cognitive science. *Minds and Machines*, 18(3), 331–348.
- Zgaljardic, D. J., Yancy, S., Levinson, J., Morales, G., & Masel, B. E. (2011). Balint's syndrome and post-acute brain injury rehabilitation: a case report. *Brain Injury*, 25(9), 909–917. <https://doi.org/10.3109/02699052.2011.585506>.
- Zipoli Caiani, S., & Ferretti, G. (2017). Semantic and pragmatic integration in vision for action. *Consciousness and Cognition*, 48, 40–54. <https://doi.org/10.1016/j.concog.2016.10.009>.
- Zipoli Caiani, S. (2014). Extending the notion of affordance. *Phenomenology and the Cognitive Sciences*, 13(2), 275–293. <https://doi.org/10.1007/s11097-013-9295-1>.
- Zipoli Caiani, S. (2018). Intensional biases in affordance perception. *Synthese*. <https://doi.org/10.1007/s11229-018-02049->.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Index

## A

Abstract, 282, 283  
Accomplish, 274, 278  
Accomplishment, 287, 288  
Achievement, 283, 288, 292  
Acting for pleasure, 287  
Action, 266, 289  
Action cascade, 266–304, 427–435  
Action categorization, 303, 425–435  
Action concept, 282–284, 287, 302  
Action frame, 289, 291, 293  
Action-related language processing, 439–441, 443, 454, 456, 457  
Action verb, 263, 266, 280–283, 288, 300, 303, 439–444, 454–457  
Action verb meaning, 266, 279, 299  
Act-tokens, 266–268, 268, 270, 282, 290, 302  
Act-tree, 263, 266–268, 270, 277, 280, 302  
Act-TTs, 266, 268, 274, 275–277, 279, 280, 289, 296, 300, 302  
Act-types, 266–268, 274, 276, 277, 279–281, 283, 286, 291, 293, 298, 300  
Additivity, 57, 58  
Adjectives, 3, 11, 12, 65–69, 71–86, 88–91, 93–95  
Affordance, 465, 467–469, 471–473  
Agency, 296, 297  
Agent, 264, 265, 267–272, 274, 277, 280–282, 284, 287–289, 291, 293–298, 301–303  
Agent-dependent interpretation, 40  
Agent roles, 303

Agent's epistemic perspective, 25, 31  
Aktionsart, 220, 221, 235, 236  
Alzheimer's, 111  
Analytic, 104, 108, 110, 113, 117, 118  
Analytical sentence, 4  
Analyticity, 103, 104, 106–109, 112–117  
Analytic/synthetic distinction, 3, 103–108, 118  
Animal, 10, 17  
Animator, 296  
Applicative, 284  
Appraisal, 286, 287, 304  
Argument augmentation, 276  
Aristotle, 3, 8  
Arten des Gegebenseins, 374  
Ascribed situation, 37  
Aspect, 3, 14, 15  
Asymmetric, 272, 277, 296  
Atomic, 114, 118  
Atomic concepts, 118  
Atomism, 104, 109, 112, 114, 116, 118, 119  
Atoms, 114  
Attenuating adverb, 439, 442–444, 448, 449, 451, 454–456  
Attitude ascriber, 37  
Attitude content, 26–29, 31–34, 39, 40  
Attitude embeddings, 27, 30  
Attitude report content, 25, 26  
Attitude reports, 30, 41  
Attitude verbs, 3, 37  
Attitudinal embedding, 25, 36, 39  
Attribute, 1–3, 9, 10, 15–17, 289, 290  
Attribute-value structure, 1, 9

Augmentation generation, 270, 271, 275–277, 284–287, 293  
 Austin, 264, 282, 291–293  
 Author, 294, 296, 301  
*average*, 65, 67, 68, 70, 71, 73, 78, 80–88, 92–96

**B**

Barsalou, 1, 3, 167–169, 173, 176, 224  
 Barsalou frames, 6–10, 127–138, 223–235, 244–248, 251–258, 288–291, 313–317, 333–343  
 Base-labeled feature structures, 224  
 Basic act, 279–282  
 Bayesian, 329–333, 336–338, 342–344  
 Bayes’ theorem, 316  
 Bearer of the attitude, 37  
 Beta desynchronization, 443, 452, 455  
 Brooks, 8  
 Brute fact, 281  
*by gerund*, 283

**C**

Carnap, 7  
 Cascade relation, 1, 3, 14, 16, 18, 19, 278, 287, 290. *See also* C-constitution  
 Cascades, 263, 266, 274, 277, 279, 281, 283–304, 412, 425–435  
   - and categorization, 264–266, 273–279, 303, 427–433  
   - and cognition, 274, 303–304, 426–431  
   - and composition, 300–302  
   - and frames, 288–291  
   - and learning, 304, 430–431  
   - and meaning, (general sense), 304  
   - and meaning (semantics), 299–302  
   - and practical knowledge, 304, 430–431  
   - and reality, 303–304  
   - and social interaction, 284–285, 428–429  
   - and speech-act theory, 291–299  
   - and verb classes, 279–288  
   - in rat psychology, 431–435  
 Categorization, 8, 9, 263, 264, 266, 273, 276, 279, 303, 329–332, 334–338, 340–344. *See also* multilevel categorization  
 Category, 9, 16, 17  
 Category learning, 329–333, 335, 338, 342, 343  
 Causal generation, 269, 288  
 Causality, 48

Causative, 109–111, 115, 116, 272, 278, 283, 286–288  
 Causative upshot, 283  
*c-by*, 277, 290  
 C-constitution, 263, 266, 277, 279, 284, 290, 291, 296–299, 301–304  
 C-implementation, 296, 297  
*C-in*, 277, 290, 292  
 Centered informational situations, 31  
 Centered situation, 31, 36, 37, 39  
 Central node, 289, 291, 301  
 Character, 29  
 Choice function, 37  
 Circumstances, 263, 264, 266, 270, 271, 276, 278, 280, 281, 292, 293, 296–300, 302, 304  
 Classical notion of compositionality, 109  
 Classical way of conceiving, 109  
 Clausal embedding, 26  
 Clausally complemented verbs, 36  
 Coercion, 55, 56, 104, 107–109, 112, 113, 117  
   - type-coercion, 112  
   - type-shifting, 112  
 Cognition, 6, 7, 10, 12, 16, 18, 19, 123–125, 136, 137, 166–167, 283, 302–304, 343, 428  
 Cognitive architecture, 103–105  
 Cognitive content, 41  
 Cognitive flexibility, 390, 391  
 Cognitive opacity, 26  
 Cognitive psychology, 1, 6, 9, 107  
 Cognitive representation, 266, 274, 275, 288, 299, 302, 412, 417, 425, 426, 431  
 Cognitive science, 103, 104, 107, 109  
 Cognitive semantics, 109  
 Cognitive structures, 389–391, 402, 406  
 Cognitive transparency, 26, 30  
 Cognitively opaque contexts, 26  
 Cognitively transparent contexts, 26  
 Collective interpretations, 58  
 Color, 121–139  
   - labels, labelling, 121–140  
 Color perception, 3, 124, 126  
 Combination, 45, 48, 55  
 Common ground, 201, 213, 215  
 Common nouns, 3, 50  
 Communication, 202, 207, 213  
 Complex representations, 107  
 Complex templates, 110  
 Compose into complex structures the classical way, 114

- Composition, 4–7, 11, 12, 15, 16, 266, 299–302  
Compositional, 5  
Compositional interpretation of attitude reports, 29  
Compositionality, 40, 106–109, 311, 312, 325, 326  
- classical compositionality, 109  
- enriched compositionality, 106. *See also* Enriched form of compositionality  
Compositionality / compositional interpretation / compositional semantics, 25, 27–31, 33–36, 38, 39  
Compound augmentation, 276, 277  
Compound generation, 275  
Computational, 105  
Computational explanation, 464  
Computational paradigm, 463–465, 468  
Computational processes, 104, 114  
Computations, 105, 114  
-linguistically-driven computations, 105  
Concepts, 1, 2, 4, 6–9, 12, 13, 16, 19, 49–55, 60–62, 106–111, 113–116, 115, 118, 121, 122, 124, 127, 128, 133–136, 138  
Conceptual, 104, 110  
-composition, 117  
-constituents, 106  
-content, 103, 116  
-Holism, 118  
-tokening, 113, 117, 118  
Conceptualization, 274, 303  
Conceptual semantics, 109  
Conceptual spaces, 1–3, 13, 17, 146, 147, 151, 152  
Configuration, 43, 46, 49, 51, 53, 60, 62  
Configurational entities, 43  
Configurational objects, 62  
Constraint(s), 181–183, 186–189, 193–196, 313–320, 323, 325, 326. *See also* Modification effect, relevant modification  
Content, 1–4, 11, 13, 29  
Content pluralism / pluralism about linguistic content, 25, 40  
Context, 201–205, 209–216  
Context-dependence, 25  
Context-dependent meaning, 25  
Contexts, 29  
Contextual knowledge, 302  
Contextually chosen, 36  
Contextually determined interpretation, 41  
Conventional generation, 269, 281  
Conversational implicature, 300  
Convex domain, 54, 56  
Convexity, 2  
Convex set, 53  
Co-temporal, 271, 273, 276, 293  
Counting, 3, 7, 11, 12, 43, 44, 57, 58, 60, 61  
Count noun, 43, 57  
Counts as, 281  
Covering law model of explanation, 467. *See also* Deductive nomological explanation  
Creation, 52  
Cresswell, 7  
Criterion predicate, 278, 283, 285, 286  
Cumulative interpretation, 43, 58, 59  
Cumulativity, 59
- D**
- Davidson, 7  
Decomposition, 2, 6, 7, 66, 110–111, 116, 118, 223–230, 244–248, 251–260, 287–294, 300. *See also* Predicate decomposition  
Decompositional frames, 224  
Decompositional frame semantics, 223, 236  
Decompositional frames (frame structures), 219, 223, 224, 236  
Decompositionalists, 113  
De dicto-readings, 26  
(De dicto-readings of) attitude reports, 26  
Default inheritance, 313  
Denotational content, 26  
Derivation(s), 181, 188, 191, 194, 196  
Diagnosticity, 312, 329, 330, 332, 333, 335, 336, 339, 341–343  
Diagonal, 30  
- of a character, 29  
Difference in substitutivity, 34  
Dimensions of comparison, 365, 366, 368, 371, 373, 375, 376, 380  
Directed motion, 240, 258  
Direct reference, 368  
Distance, 329, 330, 333, 336, 339–342, 344–347  
Distance (measure), 339, 344  
Doing too much, 287  
Dot object, 294  
Dowty, 6, 286, 288  
Dual stream model of vision, 472, 473  
Dynamical Systems Theory (DST), 464, 465, 467–471, 473  
Dynamic attribute, 246

**E**

- Economy, 201, 202, 213, 215  
 Electroencephalography (*EEG*), 439, 442, 443, 449–453, 456  
 Embodied cognition, 18, 19, 136–137, 440–457  
 Empty centered situation, 36, 37, 39  
 Empty situation, 36  
 Enactivism, 3, 19  
 Enriched semantic composition, 113  
 Entailment, 3, 116  
 Epistemic content, 26  
 Epistemic perspective, 25, 34  
 Equal partner principle, 466, 468  
 Erratic verbs, 286  
 Events, 105–107, 110–113, 114  
 Exemplar theory, 8  
 Expressivity, 201, 202, 213, 215  
 Extension, 7  
 Extensible Metagrammar (XMG), 181–197  
 Extensional, 2, 4, 5, 11  
 Extensional and attitude verbs, 40  
 Extensional complements, 39  
 Extensional verbs, 36, 39  
 Extension/extensional construction, 25–27, 30, 31  
 Extensible Metagrammar, *see* XMG  
 External anchor / external anchoring, 33  
 Extreme, 352, 355, 358

**F**

- Family resemblance, 8  
 Favor, 284, 427–429  
 Feature list approaches, 9–10, 128, 329–340  
 Features, 104, 118  
 Felicity conditions, 292  
 Fictive motion, 239–243, 246, 249–251, 254–256, 258, 259  
 Fillmore, 1, 167, 168, 173, 176, 177  
 First-order frames, 266, 289, 290, 291, 298  
 Flexibility, 391  
 Footing, 295, 296, 301, 302  
 Foot verb, 444, 447–449, 452, 453, 455, 456  
 Formal logic, 4  
 Formal semantics, 2–8, 11, 12  
 Frame analysis, 181–196, 233–235, 239–258, 289–294, 314–317, 329–342  
 Frame diagram, 289  
 Frame hypothesis, 7, 289  
 Frames, 1–3, 6–10, 13–17, 121, 122, 125, 127, 128, 130–132, 134, 135, 140, 167, 168, 172–177, 266, 287–293,

295, 298, 300, 301, 329, 330, 332–336, 338–344. *See also* frame analysis

- Frame-semantic representation, 226, 227  
 Frame semantics, 109, 173, 181–182, 188, 192, 219, 223–224, 233, 236, 299–302  
 Frame structure, 289, 291  
 Frame types, 225, 227, 229  
 Free-will, 390  
 Frege, 4, 7, 8, 27, 28, 374  
 Functional term, 46

**G**

- Gärdenfors, 2, 13, 17, 147, 353, 356, 361, 369, 370, 377, 382  
 Generative Lexicon, 6  
 German, 239–259, 284–287, 365–377  
 Goals, 390  
 Goffman, 295, 296  
 Goldman, 9, 263–277, 279–283, 286, 288, 289, 291, 295, 297, 302, 303, 426–428, 435  
 Gradability, 148, 157, 365–367, 375, 377, 379, 380, 384  
 Grounding, 55, 56  
   - in cognition, 137–138, 440

**H**

- Hand verb, 444, 447–449, 452, 453, 455  
 Heuristic, 105, 330, 336  
 Heuristic principles, 105  
 Holism, 112, 118  
 Host, 278, 283  
 Human action, 268, 274, 302–304  
 Hume, 8  
 Hungarian, 219, 220, 223, 226, 235  
 Hyperintensional interpretation, 30  
 Hyperintension/hyperintensional construction, 26  
 Hyperintensions, 28

**I**

- Illocutionary act, 291, 292  
 Imperfective, 202–209, 211, 212  
 Implicature, 3, 13, 14, 143–145, 147–149, 154, 156, 158, 159  
   - implicature space, 14, 151–158, 160–161  
 Incremental change, 221, 226–229, 234  
 Incremental progression, 233

Indeterminacy, 103, 104, 112  
 Indeterminate, 108, 112  
 Indeterminate sentences, 109, 115, 117  
 Indexical expressions, 29  
 Indices, 46–54, 56, 58–62  
 Indistinguishability, 365, 366, 368, 370, 377, 378, 380  
 Individual concepts, 4, 12, 43, 49–62  
 Inferences, 103, 104, 109, 113–119  
 Inferential relations, 114  
 Informational situations, 31, 32
 

- centered informational situations, 31, 32
- contextually chosen situation, 36
- the empty centered situation, 37, 39

 Information content, 26  
 Information states, 25, 32, 33  
 Institutional fact, 281  
 Integrated, 29  
 Integrated content, 25, 31, 32, 40  
 Integrated semantics, 27, 31–34, 36, 39, 40  
 Intensifying adverb, 439, 444, 448, 449, 454, 455  
 Intensional, 2–5, 15, 240, 243, 257, 258, 280, 286  
 Intensional construction, 26  
 Intensional generalized quantifiers, 33  
 Intensional verb, 5, 26  
 Intensions, 3, 4, 7, 8, 26, 28  
 Intentional fallacy, 107, 109, 112, 115, 116, 119  
 Intentionality, 275, 280, 286  
 Interaction, 274, 281, 300, 303, 304  
 Interpretation
 

- agent-dependent interpretation, 40
- contextually determined interpretation, 41

 Intra-modular, 105  
 Introspection, 390  
 Irreflexive, 272, 277, 296

**J**

Japanese, 284, 287

**K**

Kaplanian character, 29  
 Kaplanian monsters, 36  
 Kinds, 65, 66, 68, 74, 81, 83, 84, 87, 89, 90, 91–96  
 Knowledge-how, 19, 304, 430–431, 435

**L**

Label, 128–135, 137–139  
 Lakhota, 285, 286  
 Language, 121–127, 129, 130, 133–135, 138, 139  
 Language processing, 105  
 Learning, 2, 3, 10, 18, 19
 

- action learning, 304, 425–427, 430–435
- category learning, 276, 329–347
- reversal learning, 389–406

 Level-generation, 263–281, 283, 284, 286, 287, 290–292, 296, 297, 300, 302–304, 426–429, 433–434  
 Level of action, 280, 296, 304, 430–433  
 Lexical causatives, 103, 104  
 Lexical-conceptual, 118  
 Lexical items, 103, 104, 106  
 Lexical meaning, 5–7, 266, 283, 286, 288, 289, 292, 299–302  
 Linear mixed effect model, 446, 450  
 Linguistic cognition, 125  
 Linguistic content, 25, 40
 

- attitude report content (epistemic content / information content / cognitive content / subjective meaning), 25, 26, 41
- integrated content, 25, 31–33, 40
- truth-conditional content (denotational content / objective meaning), 25–29, 31–33, 40

 Linguistic relativity, 121–124, 127, 133  
 Locke, 8  
 Locutionary act, 291, 292  
 Logic, 8  
 Logical condition, 273  
 Logical entailment, 277  
 Logical form, 105, 106, 111, 113  
 Logical independence, 277  
 Logical relations, 4, 273  
 Logical type, 5, 6, 11

**M**

Mandarin, 285, 287  
 Manner adverb, 439, 442–444, 454–457  
 Manner augmentation, 276  
 Manner modification, 294, 295  
 Manner of motion, 240, 242, 251, 259  
 Matching procedure, 444  
 Material join operation, 46, 54  
 Maurin, 171  
 Meaning, 1–4, 6, 7, 11–17, 19, 105, 106, 108–110, 113–115, 117, 118, 263, 275, 299, 300, 302–304

- Meaning postulates, 13, 115, 116  
 Meaning representation, 2, 4, 5  
 Meanings, 106  
 Measure functions, 57, 58  
 Memory, 110, 116  
 Mental model, 18  
 Mental representations, 104, 107, 112, 119, 121–122, 127, 135, 138. *See also* Cognitive representations  
 Metagrammar, 181, 183, 191–195, 232  
 Metaphysics, 66, 67, 96  
 Method-neutral causatives, 283  
 Model-theoretic, 3, 4, 11  
 Mode of Presentation (MoP), 27, 28, 32, 33  
 Modification, 3, 16, 311–315, 317–320, 323–326
 - relevant modification, 316, 319, 321–326
 - selective modification model, 312
 - typical modification, 313, 323, 325  
 Modification effect, 313  
 Modifier, *see* Modification  
 Modifier effect, *see* Modification effect  
 Modular, 105  
 Molecularism, 118  
 Molecular representations, 114  
 Moltmann, 9  
 Montague, 3–5, 7, 11, 14  
 Montague grammar, 4  
 Morphemes, 106, 114  
 Morphology, 3, 7  
 Motor activation, 440, 454–456  
 Motor activity, 441, 443, 456  
 Motor representations, 1, 2  
 Motor system, 464  
 Movement, 3, 15  
 Multidimensional attribute spaces, 366, 370, 380, 381  
 Multidimensional scaling, 152  
 Multilevel categorization, 263, 303, 427, 430, 435
- N**  
 Names, 35  
 Natural concept, 143, 146, 147  
 Natural language metaphysics, 66, 67  
 Neuroscience, 107  
 Newen, 9  
 Non-analytic inferences, 115  
 Non-basic action, 280, 300, 303  
 Non-basic act-type, 279, 280  
 Non-extensional predicates, 52
- Nonlocal readings, 68, 76–79, 84, 93, 95  
 Non-representational account of cognition, 465
- O**  
 Objective meaning, 26  
*occasional*, 65, 68–71, 73, 78–84, 88, 89, 93, 95, 96  
 Ontology, 2, 4, 9, 12, 265, 274, 289, 295, 304  
 Optic ataxia, 463–465, 468–472  
 Overlap, 43, 45, 57, 58
- P**  
 Pacherie, 9  
 Parasite, 278, 283  
 Partee, 7, 11, 85  
 Partee temperature puzzle, 171, 173, 175  
 Path, 239–243, 245–247, 249–259  
 Peacocke, 8  
 Perception, 3, 12, 13, 111, 123–127, 129, 132, 136, 138, 139  
 Perceptual information, 121, 122, 135, 136, 140  
 Perlocutionary act, 292  
 Perner, 9  
 Person, 297, 301  
 Perspective, 201–204, 207, 208, 213–216  
 Perspective-dependence, 30, 31  
 Phatic act, 291, 292  
 Phonetic act, 291, 292  
 Pluralism, 40  
 Polysemy, 188, 189, 193, 194, 240–242, 300  
 Possible worlds, 3, 4, 7  
 Possible-worlds semantics, 3, 11  
 Practical knowledge, 304, 425–427, 430–435  
 Pragmatic, 105  
 Pragmatically, 117  
 Pragmatics, 3, 107, 143, 145, 156, 158  
 Predicates, 109–111, 114, 116  
 Predication, 5  
 Principal, 296, 301  
 Principle of compositionality, 7  
 Processes, 104, 105, 108, 114  
 Processing, 104, 113, 115
 - language, 105  
 Productive, 106  
 Productivity, 106, 119  
 Progression, 225–228, 233, 234  
 Progressive, 201–207, 209–216  
 Proper names, 29, 33

Property, 47, 49–52, 54, 56, 57, 60, 62. *See also* Features  
Proposition, 106, 108–110, 113, 114  
Propositional, 106  
Propositional content, 110  
Prosocial behavior, 18, 413, 432, 433  
Prototypes, 1, 2, 8, 16, 17, 130, 309–323, 351–355, 359  
Prototype concepts, 312  
Prototype theory, 311, 314, 317, 326. *See also* Typicality  
Pustejovsky, 6

**Q**  
Qualia, 6  
Quantification, 5  
Quantifier, 3, 6

**R**  
Radical enactivism, 463, 465, 467  
Rats, 3, 10, 18, 19, 389, 390, 392–406, 411–426, 431–435  
Reality, 282, 303  
Recanati, 9  
Record types, 3, 14, 167–175, 177  
Reference, 114, 118, 266, 267, 291, 292, 299–302  
Referent, 106, 107, 114  
Referential, 118  
Referential opacity, 26  
Referential relation, 115  
Referential transparency, 26  
Reinforcement learning, 18, 19, 415–417, 432–434  
Relevant modification, *see* Modification effect  
Representations, 6, 7, 9, 10, 13, 16–18, 105, 106, 110, 112, 114, 329–331, 333–335, 342, 343, 463, 464, 472. *See also* mental representations  
Representativeness, 360  
Representatives, 351, 353, 355, 358, 360  
Result, 285, 287  
Resultative, 224, 232, 285, 286  
Resultative predicate (resultative phrase, resultative adjective), 219, 221, 230–234  
Reversal learning, 18, 389, 391–393, 396, 397, 401–406  
Rhetic act, 291, 292  
Riddles, 44  
*rise*, 239, 240, 242, 251, 258

Role and Reference Grammar, 219, 230, 231, 233  
Roles, 263, 290, 295, 296, 298, 303, 304  
Rosch, 8  
Russian, 287

**S**  
Sameness, 380  
Scalar analysis (scalar approach), 223  
Selectional restrictions, 113  
Selective modification model, 312–315  
Semantic complexity, 111  
Semantic templates, 110, 111  
Semantic theory, 4, 11, 16  
Senses, 27, 113, 117  
Sequence, 44, 46  
Sequential, 278, 294  
Set shifting task, 10  
Set theory, 4  
Signals, 284, 296  
Similarity, 3, 17, 365–372, 374–381, 384, 385  
Similarity demonstratives, 365–368  
Simple generation, 270  
Situation, 167–173, 177  
SMM, *see* Selective modification model  
Social action, 281, 303  
Social calls, 10  
Social interaction, 284, 294, 296, 304  
Spanish, 201–204, 209, 211, 212, 214–216  
Speech act cascade, 282, 291, 292  
Speech act theory, 264  
Standards of information, 37  
States, 103, 104, 106, 110, 111  
Stative reading of dynamic verbs, 241  
Stative uses of dynamic verbs, 241  
*steigen*, 239–248, 250–259  
Stereotypes, 3, 17  
Strongest meaning hypothesis, 59  
Subconcepts, 60, 61  
Subjective meaning, 26  
Substitution-allowance, 39  
Substitution properties, 30  
Substitution/substitutivity/substitution behavior/substitution-resistance/substitution failure, 25–27, 29–31, 34, 36, 39, 40  
Subsumption, 276, 277, 290  
Sum entities, 46, 50, 57  
Sum individual concepts, 57–60  
Symbol, 118  
Symbolic, 105  
Symbolic expression, 106, 114

Symbolic representations, 105, 114  
 Symbols, 105, 114, 118  
 Syntactic gap, 117  
 Syntactic structure, 5  
 Syntax-semantics interface, 219, 230, 233  
 Synthetic, 104, 108, 113, 116–118

**T**

Tarski, 7  
 Telicity, 220, 223, 233, 234  
 Templates, 109, 111  
 Terminology, 4  
 Thematic roles, 111  
 Thematic structure, 111  
 Theory of action, 263–266, 283, 291, 302, 303  
 Theory of mind, 201, 213, 215  
 Theory theory, 8  
 Token-of-a-Type (TT), 267, 268, 277, 279, 296, 300  
 Transitive, 272, 277, 282, 283, 286, 296  
 Transitivity, 272, 290  
 Tree structure, 277  
 Truth-conditional, 11, 27, 28, 32  
 Truth-conditional and attitude content, 31  
 Truth-conditional component, 37  
 Truth-conditional content, 25, 26, 28, 29, 31–33, 40  
 Truth-conditional contribution, 25  
 Truth conditions, 2–4, 11  
 Truth-preserving substitution, 39  
 Truth value, 3, 14  
 Two-dimensional semantics, 25, 27, 29–31, 40  
 Type logic, 3  
 Types, 167–177  
 Type-shifting, 112  
 Type Theory with Records (TTR), 167–176  
 Type token, 372, 377  
 Typical, 317, 351–353, 355, 358, 359  
 Typicality, 311, 315
 

- and probability, 322, 323. *See also* Prototype theory / modification effect, typical modification

 Typicality and probability, 316, 322, 323, 325

Typical modification, *see* Modification effect

**U**

Ultrasonic vocalization, 411–413, 416–425, 431–433, 435  
 Under- or misdetermination, 28, 29  
 Unification, 289, 301, 302  
 Use (contrast with meaning and sense), 117, 118  
 Utterance meaning, 302

**V**

Van Valin & LaPolla, 284–287  
 Variation, 201–204, 209–212, 215  
 Verb-adverb-combination, 449, 454  
 Verbal particles, 219–224, 226–233, 235, 236  
 Verb class, 279, 282  
 Verb-dependent standards of information, 37  
 Verb frame, 245  
 Verb meaning, 279, 283, 290, 291, 303  
 Verbs, 3, 11, 13–16, 19
 

- basic/non-basic, 280–282
- erratic, 286
- of creation, 47
- of killing, 286, 287
- of motion, 239, 242, 259
- psychological, 111

 Visual affordances, 465  
 Visual attentional mechanism, 115  
 Vocalizations, 3, 10, 18, 19, 410–433  
 Vosgerau, 9

**W**

Wave, 51, 53, 62  
 Weights/weighted model, 329–332, 336, 340, 342, 343  
 Werning, 9  
 Wittgenstein, 8

**X**

XMG, 181–197