

Gaetano Fiorin  
Denis Delfitto

---

# Beyond Meaning: A Journey Across Language, Perception and Experience

# Perspectives in Pragmatics, Philosophy & Psychology

Volume 25

## **Editor-in-Chief**

Alessandro Capone, University of Messina, Barcelona, Italy

## **Consulting Editors**

Keith Allan, Monash University, Melbourne, Australia

Louise Cummings, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

Wayne A. Davis, Georgetown University, Washington, USA

Igor Douven, Paris-Sorbonne University, Paris, France

Yan Huang, University of Auckland, Auckland, New Zealand

Istvan Kecskes, State University of New York at Albany, Albany, USA

Franco Lo Piparo, University of Palermo, Palermo, Italy

Antonino Pennisi, University of Messina, Messina, Italy

Francesca Santuli, Università Ca' Foscari, Venezia, Italy

## **Editorial Board Members**

Noel Burton-Roberts, University of Newcastle, Newcastle, Australia

Brian Butler, University of North Carolina, Asheville, USA

Marco Carapezza, University of Palermo, Palermo, Italy

Felice Cimatti, Università della Calabria, Cosenza, Italy

Eros Corazza, Carleton University, Ottawa, Canada

Marcelo Dascal, Tel Aviv University, Tel Aviv, Israel

Michael Devitt, City University of New York, New York, USA

Frans van Eemeren, University of Amsterdam, Amsterdam, The Netherlands

Alessandra Falzone, University of Messina, Messina, Italy

Neil Feit, State University of New York, Fredonia, USA

Alessandra Giorgi, Ca' Foscari University of Venice, Venice, Italy

Larry Horn, Yale University, New Haven, USA

Klaus von Heusinger, University of Stuttgart, Stuttgart, Germany

Katarzyna Jaszczolt, University of Cambridge, Cambridge, UK

Ferenc Kiefer, Hungarian Academy of Sciences, Budapest, Hungary

Kepa Korta, ILCLI, Donostia, Spain

Ernest Lepore, Rutgers University, New Brunswick, USA

Stephen C. Levinson, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Fabrizio Macagno, New University of Lisbon, Lisbon, Portugal

Jacob L. Mey, University of Southern Denmark, Odense, Denmark

Pietro Perconti, University of Messina, Messina, Italy

Francesca Piazza, University of Palermo, Palermo, Italy

Roland Posner, Berlin Institute of Technology, Berlin, Germany

Mark Richard, Harvard University, Cambridge, USA

Nathan Salmon, University of California, Santa Barbara, USA

Stephen R. Schiffer, New York University, New York, USA

Michel Seymour, University of Montreal, Montreal, Canada

Mandy Simons, Carnegie Mellon University, Pittsburgh, USA  
Timothy Williamson, University of Oxford, Oxford, UK  
Anna Wierbizcka, Australian National University, Canberra, Australia  
Dorota Zielinska, Jesuit University of Philosophy and Education  
Ignatianum, Kraków, Poland

More information about this series at <http://www.springer.com/series/11797>

Gaetano Fiorin • Denis Delfitto

# Beyond Meaning: A Journey Across Language, Perception and Experience

 Springer

Gaetano Fiorin  
Utrecht University  
Utrecht, The Netherlands

Denis Delfitto  
University of Verona  
Verona, Italy

ISSN 2214-3807                      ISSN 2214-3815 (electronic)  
Perspectives in Pragmatics, Philosophy & Psychology  
ISBN 978-3-030-46316-8              ISBN 978-3-030-46317-5 (eBook)  
<https://doi.org/10.1007/978-3-030-46317-5>

© Springer Nature Switzerland AG 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

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

*A Linda e Paola*

*“What’s in a name? That which we call a rose  
By any other name would smell as sweet”*

*WILLIAM SHAKESPEARE, Romeo and Juliet*

# Acknowledgments

Our work on first-person reference in natural language began 15 years ago and has so far not once failed to keep us occupied and, above all, puzzled. One thing that has become clear to us is that a satisfactory understanding of the notion of self-reference in natural language cannot be provided outside the scope of a theory of linguistic meaning that expresses all its three main dimensions: its linguistic dimension—the way meaning is conveyed by linguistic forms; its referential dimension—the way meaning relates to the material conditions of the environment in which it occurs; and its cognitive dimension—the way meaning affects the cognitive lives of its users. This book was conceived as an attempt at addressing some of the main puzzles we met along the way within a coherent story about the *nature* of linguistic meaning. In fact, to us this book feels like nothing more than a necessary premise and certainly not as a conclusion. The issues of linguistic meaning addressed here span from relatively technical puzzles like the logical form associated with sentences involving implicit *de se* to broad philosophical questions concerning whether the notion of truth (and a theory of truth) constitutes a necessary and sufficient foundation for an adequate theory of linguistic meaning. However, the goal we set for ourselves here was in a sense less oriented towards a detailed discussion of these issues—to which we devoted some efforts in the past and intend to devote further attention in the future—than towards an elucidation of the preliminary questions that need be answered in order for matters of linguistic meaning to be satisfactorily investigated. In our view, these questions revolve around the cognitive foundations of the *representational* power of language and demand a precise assessment of the relation of all ingredients of linguistic meaning with pivotal aspects of cognition such as *perception* and *experience*. Whether we succeeded or failed in our enterprise of making the case for the systems of perception being closer to the language systems of interpretation than many formal linguists would be ready to concede is not for us to judge; all we can say with certainty is that for us this endeavor was worthy, if not for other reasons, at least for the many brilliant minds it has allowed us to meet all along the way. We are grateful to all of them for sharing with us their ideas, their criticism, and their encouragement. A pivotal role in the first stages of our research

was played by our mentor and friend James Higginbotham, who inspired us with his groundbreaking ideas and guided us with his sharp criticism.

Gaetano Fiorin wishes to express his gratitude to the teachers and mentors who introduced him to the fields of linguistics and semantics—among them (and beyond Denis), Lisa Cheng, Gennaro Chierchia, Norbert Corver, Giorgio Graffi, Irene Heim, Andrea Moro, Tanya Reinhart, Eric Reuland, and Frank Wijnen; to the many colleagues and friends who, throughout the years and in all sorts of ways, shared his journey—among many others, Jakub Dotlačil, Chiara Melloni, Rick Nouwen, Elena Pagliarini, Giorgos Spathas, and Maria Vender; to his colleagues and students at UCU—in particular, Elisabeth Angerer, Jocelyn Ballantyne, Floris van der Burg, Nina Köll, Francesco Maiolo, Johannes van Ophuijsen, Rosemary Orr, Chiara Robbiano, the students of the Wittgenstein reading group, Kerrewin, Merel, and Olek, and, most especially, Tijana Zakula, who patiently shared her insights of Art Historian on Malevich, Gombrich, and pictorial perspective; to Concha and her family in Granada for allowing him to spend much of the summer of 2018 writing in their beautiful patio; to his family for their endless support and affection.

Denis Delfitto fully acknowledges his heavy intellectual debts to the philosophers and linguists who first introduced him to the beauty and excruciating puzzles of the theories of meaning: Paolo Casalegno, Noam Chomsky, James Higginbotham, Giuseppe Longobardi, Tanya Reinhart, and Eric Reuland. His privilege was not so much to be allowed to share some of their extraordinary insights, it was rather to be allowed to share their sense of perspective: in science, as in a painting, it is perspective that turns flat images into alive three-dimensional objects. A special thanks goes to the Italian and Dutch friends whose unfailing friendship has been a constant source of support across the years: Pier Marco Bertinetto, Lisa Cheng, Norbert Corver, Martin Everaert, Giuseppe Longobardi, Chiara Melloni, Eric Reuland, and Maria Vender. And in more recent times, to Anne Reboul, whose astonishing knowledge of the literature in the philosophy of mind was a constant source of inspiration. Denis is also sincerely grateful to all friends and colleagues who made, in different periods of his life, his journey through the sciences of language a pleasant and unique experience: Sergey Avrutin, Alessandro Capone, Gennaro Chierchia, Hamida Demirdache, Joseph Emonds, Alessandra Giorgi, Giorgio Graffi, Maria Teresa Guasti, Manfred Krifka, Jacques Moeschler, Andrea Moro, Henk van Riemsdijk, Luigi Rizzi, Uli Sauerland, Ur Shlonsky, Alessandra Tomaselli, among many others.

George Orwell once said that writing a book is an excruciating enterprise. We agree, although, in our case, we could benefit from the unwavering support and guidance of Alessandro Capone—the director of the series *Perspectives in Pragmatics, Philosophy, Psychology*—and of our editors at Springer Helen van der Stelt, first, and Anita van der Linden-Rachmat, later. We are also grateful to two anonymous reviewers whose detailed feedback and constructive criticism have allowed us to improve the original manuscript substantially. Needless to say, all errors and omissions are ours.



# Contents

<b>1</b>	<b>Introduction</b> .....	1
<b>Part I Meaning and Objects</b>		
<b>2</b>	<b>Beyond Sense</b> .....	7
	References and Remarks .....	9
	References .....	11
<b>3</b>	<b>Meaning and Language</b> .....	13
	Meaning as Something .....	13
	Meaning as Nothing .....	15
	A Kuleshov Effect, in Language .....	17
	References and Remarks .....	19
	References .....	20
<b>4</b>	<b>Meaning and Compositionality</b> .....	21
	<b>Principle of Compositionality</b> .....	21
	References and Remarks .....	23
	References .....	23
<b>5</b>	<b>Meaning and Grammar</b> .....	25
	Finite-State Grammar .....	26
	Phrase-Structure Grammar .....	27
	Grammar and Compositionality .....	31
	References and Remarks .....	32
	References .....	33
<b>6</b>	<b>Models and Formal Languages</b> .....	35
	Models .....	36
	The Language of Models .....	37
	Language as a Mirror .....	40
	References and Remarks .....	41
	References .....	42

<b>7</b>	<b>Meaning and Functions</b> . . . . .	43
	Montague's Experiment . . . . .	43
	Proper Names . . . . .	43
	Intransitive Verbs. . . . .	44
	Declarative Sentences . . . . .	45
	Sets, Properties, and Functions . . . . .	46
	Functional Application . . . . .	48
	Montague Grammar . . . . .	51
	References and Remarks. . . . .	52
	References. . . . .	54
<b>8</b>	<b>Meaning and Relations</b> . . . . .	57
	Transitive Verbs. . . . .	57
	Relations in Model-Theory . . . . .	57
	Relations in Formal Language . . . . .	59
	Relations in Natural Language . . . . .	60
	Functions of Functions . . . . .	62
	Relations in Montague's Semantics . . . . .	66
	References and Remarks. . . . .	67
	References. . . . .	67
<b>9</b>	<b>Meaning and Quantification</b> . . . . .	69
	Determiners. . . . .	69
	Quantification in Model-Theory . . . . .	70
	Quantification in Montague Grammar . . . . .	72
	References and Remarks. . . . .	74
	References. . . . .	76
<b>10</b>	<b>Meaning and Description</b> . . . . .	79
	Definite Descriptions . . . . .	79
	Reference and Description . . . . .	80
	References and Remarks. . . . .	80
	References. . . . .	82
<b>11</b>	<b>Meaning and Possibility</b> . . . . .	83
	Truth-Values . . . . .	83
	Truth-Conditions. . . . .	85
	Possible Worlds. . . . .	86
	Propositions. . . . .	89
	The Algebra of Entailment . . . . .	91
	Possible Worlds and Compositionality . . . . .	93
	References and Remarks. . . . .	95
	References. . . . .	96

**12 Meaning and Necessity** . . . . . 97  
 Necessary and Contingent. . . . . 97  
 A Priori and A Posteriori. . . . . 99  
 De Re and De Dicto . . . . . 99  
 References and Remarks. . . . . 101  
 References. . . . . 103

**13 Meaning and Attitudes** . . . . . 105  
 Sentences Within Sentences . . . . . 105  
 Propositions Within Propositions . . . . . 106  
 Modal Bases . . . . . 109  
 Attitudes De Re and De Dicto . . . . . 110  
 References and Remarks. . . . . 111  
 References. . . . . 113

**14 Natural Language Metaphysics**. . . . . 115  
 Quantity and Quality. . . . . 115  
 Natural Language Metaphysics. . . . . 116  
 Quantity Before Quality . . . . . 118  
 Time . . . . . 118  
 Events . . . . . 119  
 Natural Language Metaphysics Versus Metaphysics Proper . . . . . 121  
 References and Remarks. . . . . 122  
 References. . . . . 124

**Part II Meaning and Subjects**

**15 Beyond Form** . . . . . 127  
 World and Mind . . . . . 128

**16 Meaning and World** . . . . . 129  
 Semantic Externalism . . . . . 129  
 Private Language (and Why You Shouldn't Try It). . . . . 130  
 The Twin Earth Thought Experiment . . . . . 132  
 Meaning Ain't in the Head . . . . . 134  
 References and Remarks. . . . . 134  
 References. . . . . 136

**17 Meaning and Actions** . . . . . 137  
 Meaning and Rules . . . . . 137  
 Meaning as Use. . . . . 139  
 Speech Acts. . . . . 140  
 Meaning as Use Means No Meaning . . . . . 141  
 Meaning and Linguistic Theory . . . . . 141  
 Language and Learning. . . . . 142

Meaning and Community . . . . . 144

References and Remarks . . . . . 145

References . . . . . 146

**18 Meaning and Reality . . . . . 147**

Things as They Are and Things as We Talk About Them . . . . . 147

Things that Count and Things that Don't . . . . . 149

References and Remarks . . . . . 151

References . . . . . 152

**19 Meaning and Ideas . . . . . 153**

A World of Ideas . . . . . 154

Grasping Ideas . . . . . 155

References and Remarks . . . . . 157

References . . . . . 157

**20 Meaning and Subjectivity . . . . . 159**

Private Language (And Why People Use It Anyways) . . . . . 159

Ouch! . . . . . 160

Subjectivity . . . . . 160

References and Remarks . . . . . 161

References . . . . . 162

**21 Meaning and Indexicality . . . . . 163**

Indexicals . . . . . 163

Indexicality and Reference . . . . . 164

Indexicals as Descriptions . . . . . 166

I am not Here Now . . . . . 169

Find Yourself . . . . . 171

References and Remarks . . . . . 173

References . . . . . 175

**22 Meaning and Context . . . . . 177**

Perspective . . . . . 177

Context . . . . . 178

Kaplan's Logic of Indexicals . . . . . 180

Content and Character . . . . . 182

References and Remarks . . . . . 183

References . . . . . 184

**23 Meaning and Causality . . . . . 185**

Two Stars, One Planet . . . . . 186

Names as Descriptions . . . . . 187

The Character of Names . . . . . 188

Character Generalized . . . . . 190

Necessary A Posteriori and Contingent A Priori . . . . . 190

References and Remarks . . . . . 191

References . . . . . 193

<b>24</b>	<b>Meaning and Acquaintance</b> . . . . .	195
	De Re and De Se . . . . .	195
	The Beholder’s Share . . . . .	197
	Acquaintance and Description . . . . .	197
	Acquaintance and Self-Acquaintance . . . . .	199
	Character Revised . . . . .	200
	References and Remarks . . . . .	201
	References . . . . .	202
<b>25</b>	<b>Attitudes De Se</b> . . . . .	203
	Attitudes De Re and Attitudes De Se . . . . .	203
	Value and Logic of Attitudes De Se . . . . .	206
	The Grammar of Attitudes De Se . . . . .	207
	Logophoricity . . . . .	208
	References and Remarks . . . . .	210
	References . . . . .	211
<b>26</b>	<b>Worlds and Centers</b> . . . . .	213
	Limits of Hintikka’s Semantics . . . . .	213
	Centered Possible Worlds . . . . .	214
	Centered Propositions . . . . .	216
	Hintikka’s Semantics Revised . . . . .	216
	The Semantics of Attitudes De Se . . . . .	217
	Irrational Kazimir . . . . .	218
	Acquaintance Generalized . . . . .	221
	References and Remarks . . . . .	222
	References . . . . .	223
<b>27</b>	<b>Meaning and Epistemic Subjectivity</b> . . . . .	225
<b>28</b>	<b>Implicit De Se</b> . . . . .	229
	Error Through Misidentification . . . . .	229
	Immunity to Error Through Misidentification . . . . .	231
	Introspection . . . . .	231
	Implicit De Se Propositional Attitudes . . . . .	233
	Evidentiality . . . . .	235
	References and Remarks . . . . .	236
	References . . . . .	238
<b>29</b>	<b>Meaning and Ontological Subjectivity</b> . . . . .	239
	Introspection and Ontological Subjectivity . . . . .	239
	Reference and Error . . . . .	240
	Introspection and Identification . . . . .	241
	“I” as a Non-referring Term . . . . .	242
	Cogito Ergo Sum . . . . .	244
	“I” as Expletive . . . . .	246
	References and Remarks . . . . .	250
	References . . . . .	252

**30 Meaning and Mind** . . . . . 255

    Weak Semantic Externalism . . . . . 257

    Semantic Internalism . . . . . 257

    Weak Semantic Internalism . . . . . 260

    Sense and Reference . . . . . 262

    References and Remarks . . . . . 264

    References . . . . . 266

**Part III Meaning and Perception**

**31 Beyond the Cartesian Divide** . . . . . 271

    Meaning as Interface . . . . . 271

    World, Mind, and Perception . . . . . 272

**32 World, Mind, and Perception (The Incorrect View)** . . . . . 273

    The Image View of Perception . . . . . 273

    Problems with the Image View of Perception . . . . . 276

    References and Remarks . . . . . 277

    References . . . . . 277

**33 World, Mind, and Perception (The Correct View)** . . . . . 279

    Perceiving Features . . . . . 279

    Feature-Maps and the Binding Problem . . . . . 283

    Perceiving Objects . . . . . 285

    Perceiving Propositions . . . . . 288

    Veridicality vs. Narcissism . . . . . 290

    Conscious Perception and Sensorial Experience . . . . . 291

    The Scientific Model of Perception . . . . . 294

    A Note on Sensorial Experience . . . . . 296

    References and Remarks . . . . . 297

    References . . . . . 301

**34 Meaning as Perception (Logic)** . . . . . 303

    Meaning as Perception . . . . . 303

    Logic of Meaning and Logic of Perception . . . . . 304

    Objects and Properties . . . . . 304

    Propositions and Functional Application . . . . . 304

    Reference and Representation . . . . . 305

    Compositionality and Exaptation . . . . . 305

    References and Remarks . . . . . 306

    References . . . . . 307

**35 Meaning as Perception (Metaphysics)** . . . . . 309

    Reference as Tracking . . . . . 310

    Entertaining Propositions and Expressing Propositional Attitudes . . . . . 313

    Conscious Experience as Perceptual Experience . . . . . 314

    References and Remarks . . . . . 319

<b>36 Beyond Reason</b> .....	321
Meaning and Nature .....	322
References and Remarks .....	323
References .....	325
<b>Author Index</b> .....	327
<b>General Index</b> .....	331

## About the Authors

**Gaetano Fiorin** is Assistant Professor of Linguistics and Logic at University College Utrecht and a researcher affiliated to the Utrecht Institute of Linguistics (UiL-OTS). He graduated in 2010 with a thesis on the semantic competence of children with developmental dyslexia. His research work spans from theoretical linguistics—particularly the semantics of the first person and its implications for the syntax–semantics interface and the metaphysics of meaning—to experimental linguistics—especially developmental dyslexia and multilingualism.

**Denis Delfitto** is Professor of Linguistics at the University of Verona. He graduated at the Scuola Normale in Pisa and had many research and teaching experiences in a number of American and European universities. His publications—in international scientific journals and peer-reviewed volumes—span from the syntax of reference and quantification in natural language—including issues of comparative syntax and semantics—to issues of language change and language impairment. His core interest is the investigation of the human capacity for language from a broad cognitive and philosophical perspective.



# Chapter 1

## Introduction



What is meaning? Natural languages—idioms such as English and Cantonese, Zulu and Amharic, Basque and Nicaraguan Sign Language—allow their speakers to convey meaning and transmit meaning to one another. But what is meaning exactly? What is this thing that words convey and speakers communicate? Few questions are as elusive as this. Yet, few features are as essential to who we are and what we do as human beings as the capacity to convey meaning through language. The goal of this book is to investigate such capacity.

Our investigation will proceed in three consecutive steps, each corresponding to one of the three parts in which the book is divided. In part I, “Meaning and Objects”, we will present one of the most recent theoretical approaches to linguistic meaning—the theory known as *Montague grammar*, whose foundations were set by the mathematician and logician Richard Montague in the early 1970s. Montague grammar is, without doubt, the most sophisticated theory of linguistic meaning available to this day. Its value resides in its capacity to explain two central aspects of linguistic meaning. The first is its productivity. Natural languages allow speakers to rely on a finite number of simple meanings—those expressed by individual words—to produce an infinite number of complex meanings. This is done by combining individual words into larger grammatical structures—phrases and sentences—of potentially unbound complexity. Montague grammar provides us with a general algorithm to systematically derive complex meanings from simple ones in harmony with the grammatical structures that hold linguistic expressions together. The second virtue of Montague grammar lies in its ability to account for the logical properties of linguistic meaning—the fact, known since Aristotle, that the different expressions of a language stand in a network of logical relations. Montague grammar allows us to derive these relations as the automatic outcome of a restricted set of general principles. Since Montague’s first formulation, the theory has flourished into a rich independent domain of research within theoretical linguistics. It has been applied successfully to a great number of linguistic phenomena in a variety of different languages. Yet, despite its success, it also raises a fundamental question. What

determines its primitive notions and fundamental principles? Montague grammar allows us to identify the essential shape of meaning—the properties it must have to participate in the framework of combinatorial and logical relations that characterize language. But what motivates such properties?

This is the question we address in part II—“Meaning and Subjects”—by considering two opposing hypotheses. The first is the hypothesis known as *semantic externalism*. This hypothesis claims that the roots of linguistic meaning are to be found in the external world of natural things. The second hypothesis is known as *semantic internalism*. This hypothesis claims that the roots of linguistic meaning lie in the inner realm of the mind. Both views have played an important role in the history of philosophical and linguistic thought—since Plato and Aristotle. As we will see, both views come with advantages but also limitations. Semantic externalism, on the one hand, captures the inextricable link that bounds meaning to the material environment in which it occurs but fails to account for the cognitive value it has for the speakers who entertain it. Semantic internalism, on the other hand, succeeds in capturing the cognitive significance of linguistic meaning but fails to explain how something private, such as thought, can participate in the public sphere of a community of speakers. In a nutshell, semantic externalism explains the communicability of meaning whereas semantic internalism explains its intelligibility. No theory, however, offers a comprehensive understanding of both dimensions.

In part III, “Meaning and Perception”, we will try to resolve the tension between these two dimensions by exploring a third hypothesis—the hypothesis of meaning as perception. According to this hypothesis, the roots of linguistic meaning are to be found neither in the external world of material objects nor in the inner realm of the mind but, in fact, at the interface between the two—that is, in *perception*. The recent developments in the study of the neurophysiology and cognition of perception demonstrate that the ability to obtain and organize environmental information relies on a restricted set of essential notions and organizing principles. These notions and principles, we will observe, are remarkably similar, in both form and substance, to those we find at the core of the way we package information when expressing it in natural language. We will use these observations to support the view that the way we talk about things is neither a function of the way things are nor a function of the way speakers think about them. It is, rather, a function of the way things are perceived.

Two notes of caution are in order before proceeding. Firstly, this is a book about natural language and natural language meaning. It is not a book in mathematical logic, philosophy, or cognitive science. Even though the book relies on a number of insights coming from the study of formal languages, the philosophy of language, and the neuro-cognition of sensory systems, it does so with the explicit purpose of making an exquisitely linguistic point, which concerns the nature of language and its meaning.

Secondly, the book is written with the explicit purpose of being accessible by a diverse audience. The book requires no background knowledge of the topic at stake. All it requires is knowledge of English and the good will to follow us throughout the many steps our exploration will demand. As such, however, the book is also incomplete, and necessarily so, as it focuses on some aspects of the topic at stake while

overlooking others. Paying attention to everything that has been said and written on the notion of meaning would be simply impossible. It would also endanger our attempt at providing a coherent story and articulating a coherent hypothesis. To partially compensate, all major chapters conclude with a section entitled “References and Remarks”, where the interested reader can find references to the literature relevant to the topics discussed and some further indications about their major ramifications.

**Part I**  
**Meaning and Objects**

## Chapter 2



# Beyond Sense

This book is built on the opinion that we are parts of a living world.

Gregory Bateson, *Mind and Nature*

In the Summer of 1913, the poet Aleksei Kruchenykh, the musician Mikhail Matyushin, and the painter Kazimir Malevich joined forces in organizing the “First All-Russian Congress of the Poets of the Future”, which took place not far from St. Petersburg, then Petrograd. Despite being versed in different arts, the three artists shared a common goal. They wanted to free art from the dogmas of tradition. According to them, the canon of rules and conventions that was then taught in art schools had transformed art into a mere exercise in style. By and large, the young artists were expected to replicate mechanically the great models of the past. The three Russian artists found this unacceptable. For them, real art was one thing and one thing only: *creation*. The goal of their congress was, therefore, to explore radically new means of artistic expression—means that would allow artists to truly fulfill their creative imperative.

During the congress, Aleksei Kruchenykh presented as an example a new form of poetry for which he had invented the name *Zaum*. “Zaum” is a neologism obtained from the composition of the two Russian morphemes “za”, meaning “beyond”, and “ym”, meaning “sense”. It is usually translated in English as “beyond sense”. *Zaum* is a very peculiar form of poetic experimentation, based on the idea that the true role of the poet is that of creating language, not merely using it. Instead of adopting an existing language, such as Russian, *Zaum* poets made their poetry out of raw consonants and vowels, sounds and syllables, assembled together in the most original ways. For them each new poem was the birth of a new language.

In the years that followed the congress, Malevich developed similar ideas in the area of painting. Until then, the main paradigm in painting had been that of naturalism, art whose goal is to reproduce objects realistically. In Malevich’s book, however, naturalism was not art at all. For him, there was nothing creative in mechanically reproducing something that already exists. In reaction to this tradition, Malevich published in the December of 1915 the artistic manifesto “From Cubism to Suprematism in Art, to New Realism and Painting, to Absolute Creation” and with

it started the artistic movement known as *Suprematism*. To be truly creative, Malevich states in his manifesto, painters must do with images what Zaum poets did with language. Zaum poets tore language down to its barest pieces so that they could use them as the building blocks of a new, truly creative, poetic language. Painters must now find the simplest constituting elements of the image so that they can use them as the starting vocabulary of a new visual language. But what are the bare constituting elements of the image? Zaum poets had identified the raw ingredients of language by looking “beyond sense”, that is, beyond the mere descriptive function of language. The same, Malevich submits in his manifesto, must be done for the image. Painters must revert their attention from the *meaning* of the image—what the image depicts—to its pure *form*—what remains when the image is stripped of everything that is meaningful. When this is done, we find that all images are the products of two essential ingredients: primary color and geometric form. These are the irreducible atoms of the image and the starting vocabulary of Malevich’s new artistic language.

This book draws inspiration from the radical approach of Zaum poets and Malevich’s Suprematism to tackle yet a different problem, that of *human natural language* and its *meaning*. By human natural language, we mean the family of languages human beings learn and speak around the world as one of their primary means of communication. It includes people’s native languages—the one or more languages they grow up with from birth—as well as those they master later in life. It also includes all the dialects, registers, and styles in which languages varies across space and time, all the different stages of their learning, their pathologies, and all their different modes of articulation—speaking, writing, signing, and so forth.

Few would disagree that one of the primary functions of natural language is to convey meaning. But what is meaning? Observe the black symbols now lying in lines on the white page in front of you. The reason why they are of interest to you is that they carry meaning. True, the symbols themselves may have some aesthetic value of their own, as typographers know very well. Yet, that is not the reason why you are looking at them now. The symbols deserve your attention primarily because they are vehicles of meaning. They communicate something to you. So, what is this thing that, at this very moment, is being conveyed from us to you by means of a sequence of black signs on a white page?

When faced with this question for the first time, many react by saying that meaning is *information*. This, however, is not really an answer. It is simply using a different name for the same thing. Calling it information, instead of meaning, does not help us addressing the real question: What sort of object is it? Is it concrete or abstract? Is it psychological or natural?

To be sure, the question of what meaning is is a very puzzling one. At first, it is even difficult to warp our head around the idea that meaning is something we can ask questions about. After all—one may react—we all know what meaning is, because we all *grasp* it. How can we understand what a word or a sentence means, if not in reason of the fact that we know its meaning? The very fact that you are now understanding the words you are reading shows that you know their meaning. If you did not know their meaning, you would not understand them. Yet, there is a

substantial difference between grasping the meaning of a linguistic expression and knowing what its meaning is. As an analogy, consider the case of perception. Through our eyes, we see colors. Colors provide us with information about the world around us. Yet, the fact that we see colors does not tell us what colors are. It does not provide us any knowledge of the physical facts that produce colors. It does not inform us about the physiological and cognitive processes that occur in our eyes and brains when we perceive colors. The same goes for language. The fact that we grasp the meaning of what we say, hear, write, or read tells us nothing about what meaning is or the process that allow us to grasp it.

On closer reflection, it is actually quite astounding that we know so little about meaning. The expressive power of natural language is quite spectacular and certainly fundamental to our lives as human beings. Through language, we talk about people, shapes, dimensions, quantities, properties, relations, colors, organisms, mechanisms, instruments, images, numbers, events, causes, knowledge, beliefs, desires, hopes, expectations, possibilities, theories, hypotheses, predictions, matter, society, structure, space, time, history, ideas, art, politics, medicine, law, peace, victory, success, loss, fear, gender, taste, soul, religion, spirit, dreams, and nightmares. We make statements, assertions, declarations, questions, exclamations, objections, literature, poetry, movies, and philosophy. We give answers, replies, orders, instructions, and directions. We even use language to talk about language itself, as we do in this book. Meaning is an essential ingredient of our lives. We are constantly surrounded by it, from the moment we wake up to the moment we fall asleep and, sometimes, even in our dreams. We would be very different creatures if we did not have the ability to learn and use meaningful language. Meaning is without doubt one of the most fundamental constitutive elements of human nature.

So, what is meaning? In our first attempt at addressing this question, we will endorse a strategy similar to that adopted by Malevich and the Zaum poets. Instead of looking at the content of meaning, we will look at its pure form—what remains of it when we have stripped meaning of everything that is meaningful. In a way, we will look beyond the sense of sense itself.

## References and Remarks

The notion of information, which, as we saw, is often associated with the notion of meaning, is a very ubiquitous one. It occurs in different contexts, formal and informal, with different connotations and is used in a number of different disciplines, from physics to psychology, with different purposes. In the context of our discussion, it is worth mentioning the pioneering work of the mathematician Claude Shannon. His [1948](#) article “A Mathematical Theory of Communication” offered for the first time an explicit quantitative framework for measuring the amount of information associated with a signal and has since been regarded as the foundation stone of modern information theory. The theory developed by Shannon in his article defines the quantity of information conveyed by a particular message as inversely

proportional to the predictability of that message. The core idea is that the more a message meets the expectations of its receiver the more obvious and, therefore, less informative it is. Conversely, the less a message meets the expectations of its receiver the less obvious and, therefore, more informative it is. In the context of the present discussion, it is important to point out that, in Shannon's mind, information is distinct from meaning. Whereas information concerns the expectation of a certain message occurring in the context of a set of possible messages, meaning is about the content of the message. The literature on the topic is truly vast. A technical, yet accessible introduction to Shannon's theory of information, its functioning, implications, and more recently developments is offered in the first chapters of C. Randy Gallistel and Adam Philip King's book *Memory and the Computational Brain: Why Cognitive Science will Transform Neuroscience* (Gallistel and King 2009).

Another common answer to the question of what meaning is identifies the meaning of words with their definition. This is, in effect, the approach to meaning we find in dictionaries. When we wish to learn the meaning of a new word, we consult a dictionary and find its definition. The use of the notion of meaning definitions in formulating a theory of meaning was heavily criticized by Jerry Fodor, Merrill Garrett, Edward Walker, and Cornelia Parkes in their classic article "Against Definitions" (Fodor et al. 1980). The article presents a number of arguments against the notion of meaning definitions. We can appreciate how the notion fails to provide us with a satisfactory answer to the question of meaning by briefly illustrating one of these arguments. Consider the sentence "Mary killed John". Whatever it is that speakers of English interpret as the meaning of this sentence, it must be such that it enables them to draw the inference that "John died". What explains the logical relation between the two sentences? It is tempting to explain the logical inference by simply replacing the verb "kill" with its definition. If saying "kill" is equivalent to saying "cause to die", then we have an explicit explanation of why "Mary killed John" entails that "John died". However, understanding the meaning of "kill" on the basis of its definition "cause to die" is not enough. To complete the explanation, we must define what "cause" means and what "die" means. That is, we must provide new definitions for the terms "cause" and "die". Such definitions would introduce other terms which, in turn, would need to be defined on the basis of other new terms, which in turn would need to be defined on the basis of other new terms, ultimately drawing us to a potentially infinite recursive loop. On the basis of this and other arguments, Fodor and his colleagues show that the notion of meaning definition cannot play any role in a theory of linguistic meaning. Notably, a potential escape from the argument we have just reviewed consists in identifying a core of primitive linguistic terms whose meaning can be defined on purely empirical grounds, without reference to other linguistic terms. We will discuss this approach in Chap. 22 of part II.



## References

- Fodor, J. A., Garrett, M. F., Walker, E. C. T., & Parkes, C. H. (1980). Against definitions. *Cognition*, 8, 263–367.
- Gallistel, C. R., & King, A. P. (2009). *Why cognitive science will transform neuroscience*. Oxford: Wiley-Blackwell.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal*, 27, 379–423.

## Chapter 3

# Meaning and Language



The very first step in our inquiry will not be a step forward but, in fact, a step backward. Any quest into the nature of meaning must in fact begin with a fundamental question: Is there such a thing as meaning? Is meaning something that we can regard as a genuine object of study? Or are we chasing an illusion, an artifact of something more fundamental? There are two possible views on this issue. One is that there is such a thing as meaning. We will refer to it as the view of *meaning as something*. The other is that there is no such thing as meaning. We will refer to it as the view of *meaning as nothing*.

### Meaning as Something

In the western philosophical and linguistic tradition, the view of meaning as something amply predates the view of meaning as nothing. Its most paradigmatic incarnation is the so-called theory of *meaning as reference*. The theory is canonically associated with the early Christian philosopher Augustine of Hippo, although it can be traced back to as far as Plato. Its fundamental idea is simple and intuitive: The meaning of a linguistic expression is *the object it refers to*. This idea is most readily exemplified by expressions such as *proper names*. Take, for instance, the proper name “Kazimir Malevich”. According to the theory of meaning as reference, its meaning is the object the name refers to—that is, the Russian abstract painter who was born in 1879 and died in 1935 and who founded the artistic movement known as Suprematism.

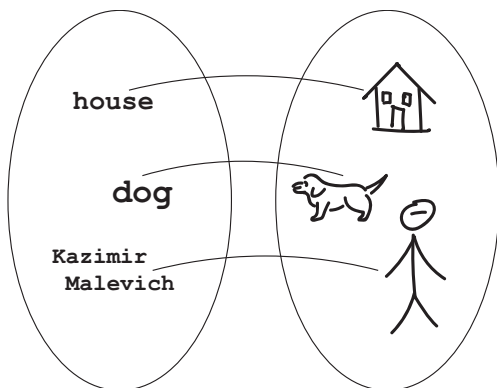
As we will see shortly, this theory runs into a number of problems. One of them, as some readers may have already suspected, is how it can be extended to words other than names—verbs, adjectives, articles, prepositions, and so forth. A further

problem is due to those expressions that may refer to different objects in different occasions. For example, the nominal expression “the king of France” may pick different referents when used to describe different periods of time. In fact, it may even pick no referent at all if used in the current times, given that France is no longer a monarchy. Does this mean that the expression “the king of France” has not one but a variety of different potential meanings, including no meaning at all? A related problem is offered by those expressions that pick the same referent in different manners. The two nominal expressions “the honorable gentlemen” and “the idiot who lives next door” may very well be used to refer to the same person, yet using one rather than the other communicates significantly more than a bare reference.

We will return to these and many other related problems later in the book. For the time being, it is useful to briefly consider the logical structure that underlies the theory of meaning as reference. The theory makes an essential distinction between two domains of objects, *linguistic expressions*, on the one hand, and *objects of reference*, on the other, as schematically represented in Fig. 3.1. Linguistic expressions, such as “house”, “dog”, and “Kazimir Malevich” are the objects that *have* meaning. Objects of reference—or, simply, *referents*—are the objects that *are* meaning. The relation that holds between the two domains is one of *reference*. The theory of meaning as reference qualifies as a theory of meaning as something because it maintains that meaning is an altogether different thing from the words that convey it. Language, under this view, is a means to point to a domain of things that would exist even if there were no words to describe it. The task of the linguist who is interested in uncovering the nature of linguistic meaning is, accordingly, that of describing such domain of things—the objects language is about.

We can appreciate already at this early stage how the theory of meaning as reference is deeply intertwined with the notion of *truth*. According to this theory, the meaningfulness of a linguistic expression resides in its capacity to refer to the object it is about. How good a linguistic expression is in performing such function depends, therefore, on how truthful it is in describing the object it is about.

**Fig. 3.1** The figure shows the distinction between two classes objects, *linguistic expressions*, to the left, and *objects of reference* or *referents*, to the right. The lines that connect the objects in the two sets represent the relation of *reference*



## Meaning as Nothing

It was only at the beginning of the twentieth century that the view of meaning as reference was seriously challenged, for the first time, by the linguist Ferdinand de Saussure. In his renowned lectures on general linguistics—delivered at the University of Geneva between 1906 and 1911—Saussure presented a radically new view of linguistic meaning whereby the words we speak have no meaning of their own; meaning is, rather, an *emergent property* of the grammatical relations that words entertain with one another.

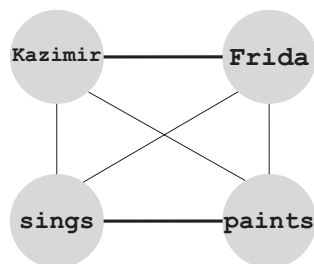
To appreciate Saussure’s paradigm, it is useful to first clarify the notion of emergent property. An emergent property is a property of a system that is not shared by its constituting elements. A classic example is temperature. Temperature is a property of physical structures made of many particles. It is determined by the speed at which the particles in the structure are moving and hitting each other. The temperature of a glass of water, for example, depends on how fast its oxygen and hydrogen particles are moving. If they are moving fast, the water is hot. If they are moving slowly, the water is cold. Crucially, temperature is a property of the system that is not shared by the individual elements that constitute it. Water has temperature. The individual particles in it, taken in isolation, do not.

Saussure’s core idea is that meaning is like temperature. It is not a property of the individual expressions of a language. It is rather a property of the relations that words entertain with each other in the larger structure of a language.

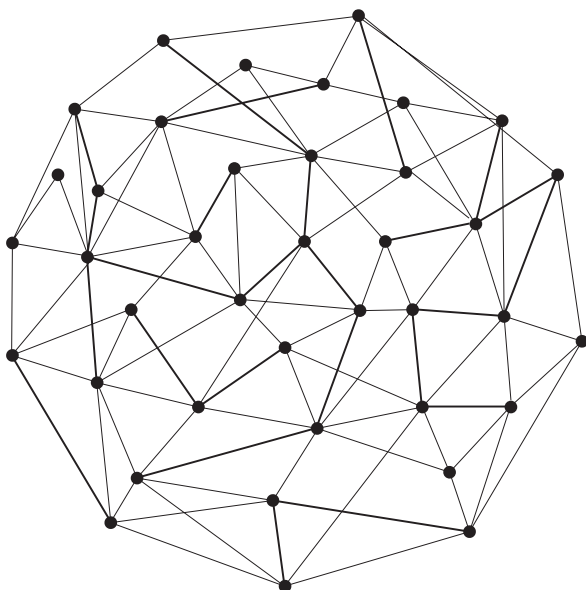
What are these relations then? For Saussure, words are held together by *grammatical* relations and grammatical relations are of two sorts: *syntagmatic* and *paradigmatic*. A syntagmatic relation is the relation that exists between two words that can occur together, one after the other, as the relation between the proper name “Kazimir” and the verb “paints” or between the proper name “Frida” and the verb “sings”. We say that these pairs of words stand in a syntagmatic relation because they can combine together, in ordered sequences, to form larger grammatical constructions. “Kazimir paints” and “Frida sings” are both legitimate grammatical constructions of English. Conversely, “paints” and “sings” do not stand in a syntagmatic relation because, when combined together, no matter in which order, they do not form a grammatical construction. “Paints sings” and “sings paints” are not grammatically acceptable constructions of English.

A paradigmatic relation is the relation that exists between words that can perform the same grammatical function and, therefore, can replace one another in the same grammatical constructions. We find paradigmatic relations between the proper names “Kazimir” and “Frida” and between the verbs “paints” and “sings”. We say that these expressions are in a paradigmatic relation because they can replace each other in the same construction without affecting the grammaticality of the construction. Take the sentence “Kazimir paints”. It is a grammatical construction of English. Now, if we replace “Kazimir” with “Frida” we still obtain a grammatical construction, “Frida paints”. This is because “Kazimir” and “Frida” are in a paradigmatic relation. Conversely, if we replace “Kazimir” with “sings”, we obtain an ungram-

**Fig. 3.2** The figure shows the grammatical relations between the four English words “Kazimir”, “Frida”, “paints”, and “sing”. The thinner lines connect the words that stand in a syntagmatic relation. The thicker lines the words that stand in a paradigmatic relation



**Fig. 3.3** A network of objects, represented by dots, and relations, represented by connecting lines. If we take the objects to be words and the relations to be grammatical relations, the figure can be taken to represent the structure of a language



grammatical combination, “sings paints”. This is because “Kazimir” and “sings” do not stand in a paradigmatic relation and, therefore, cannot replace each other in the same grammatical construction.

The system of grammatical relations of a language can be represented in the form of a *network*. We can, for example, represent the relations between the words “Kazimir”, “Frida”, “paints”, and “sing” as in Fig. 3.2.

Ideally, if we could zoom out and look at the whole system of grammatical relations of English—or, for the matter, any other language—we would obtain a larger network, roughly looking like in Fig. 3.3 (only much larger).

According to Saussure, meaning is an emergent property of networks such as the one in Fig. 3.3. It is only when we look at language as an integrated system of grammatical relations that we can characterize the notion of meaning. Individual words do not have a meaning of their own in the same way as individual particles of oxygens and hydrogen do not have a temperature of their own. Words have meaning

only and exclusively in reason of the relation they bear to the rest of the structure to which they belong. This is the reason why Saussure's theory of natural language came to be known as *structuralism*. According to this theory, the properties of a linguistic system are not a function of the individual elements that compose it but, rather, of the overall structure that holds them together.

## A Kuleshov Effect, in Language

Which view is right? Truth be told, none. Both views run into serious problems. The view of meaning as reference completely disregards the fact that grammatical relations do play a significant role in determining meaning. Sentences are not just unstructured lists of words, each with their own independent meanings. Words are put together on the basis of precise grammatical relations to form structures of higher and higher complexity, which, in turn, articulate meanings of higher and higher complexity. The view of meaning as reference provides us with individual meanings for individual words but fails to tell us how these different meanings interact with one another when they occur together in larger grammatical structures. Not surprisingly, the theory of meaning as reference works well with words such as proper names but scores very poorly with other classes of words. Surely, it is easy to identify the meaning of a proper name with its reference. Natural language sentences, however, are not unstructured lists of names. They also comprise adjectives, verbs, adverbs, articles, conjunctions, interjections, quantifiers, classifiers, intensifiers, pronouns, demonstratives, evidentials, prepositions, postpositions—to mention only a few of the grammatical classes that are attested across the natural languages of the world. It is difficult to see how the meaning of an article—such as “the”—or a preposition—such as “of”—can be characterized in terms of the object it refers to.

The structuralist view scores certainly better in this respect. Given the central role it assigns to grammatical relations, it is better suited to characterize the meaning of words such as articles and prepositions in terms of the grammatical function they perform. However, it also has its own limitations. Grammatical relations are certainly an important ingredient of meaning, but not the only one. Not all meaning, in fact, can be predicted from grammatical structure alone. Consider, as an example, sentence (1).

(1) The gostak distims the dosh

What does this sentence mean? Well, the sentence seems to be about a certain thing—“the gostak”—doing something—“distimming”—to another thing—“the dosh”. We understand at least part of the meaning of the sentence because we are able to recover the grammatical relations between the words in it. We understand that “gostak” and “dosh” are nouns and that “distims” is a verb. We also understand that “the gostak” is the subject of the sentence—the one doing the “distimming”—and “the dosh” is the object—the one undergoing the “distimming”. Yet, there is

also a part of the meaning of the sentence that we cannot recover from its grammatical structure alone. To fully understand the meaning of the sentence, we need to know what sort of things “gostaks” and “doshes” are and what sort of action “distingging” is. This information is not something we can mechanically recover from the structural relations between the words in the sentence. The lesson we learn is that individual words and the grammatical structure that holds them together are both essential ingredients of natural language meaning. The meaning of a sentence is determined by the individual meanings of the individual words that occur in it as well as by the grammatical relations that hold these words together. The view of meaning as reference and the structuralist view are both guilty of paying attention to only one of these two aspects, while disregarding the other.

To better appreciate this point, we can rely on an interesting parallel with cinematography and, in particular, with the effect known as the *Kuleshov effect*. In a famous illustration of this effect, the director Alfred Hitchcock invites us to consider a sequence of three short movie clips. The first clip is a close caption of an old man’s face (in fact, of Hitchcock himself) staring at something with a serious expression. The second clip presents a young mother tenderly holding her little child. In the third clip, we see the old man’s serious look turning into a smile. What does the spectator think of this man after having watched the whole sequence, Hitchcock asks? Most likely that he is a kind, sympathetic, old gentleman. He saw a mother holding her child tenderly and could not hold a benevolent smile. However, suppose that now we keep the first and the last clips in the sequence but replace the second—the one with the mother and the child—with a shot of young woman sunbathing in a bikini. What does the spectator think now of the old man in the last clip? To quote Hitchcock himself, the spectator is likely to think of him as “a dirty old man”, no longer “the benign gentleman who loves babies”. Through this simple comparison, we see that one and the same clip—the last clip in the sequence—is attributed two very different, in fact, contrasting meanings in the contexts of the two different sequences. Although in both contexts the clip portrays a man turning a grudge into a smile, in one case, the spectator interprets it as depicting a benevolent man, in the other, as depicting a dirty old man. The lesson cinematographers draw from this experiment is that movies tell their stories in two ways. Firstly, through the content of the individual clips that constitute them. Secondly, through the way these individual clips are put together in the overall sequence of the movie.

The lesson we can draw as linguists is that sentences are like movies. The ultimate significance of a sentence depends both on the meaning of the individual words that occur in it as well as on the way they are combined together within the sentence. This important insight was first formulated, at the beginning of twentieth century, by the philosopher and mathematician Gottlob Frege, who was one of the first thinkers in modern philosophy to embark on a systematic inquiry into the logical structure of language and meaning. His formulation of this insight is known as the *principle of compositionality*. It is to the discussion of this principle that we turn in the following chapter.

## References and Remarks

The view of meaning as reference is canonically attributed to Saint Augustine, who was the first to formulate an explicit ontological distinction between the signs of language and the meaning they convey. This point is made in book I of his *Confessions* together with a number of other remarkably interesting observations about language. This view, however, was not completely new to philosophers. It is, for example, at the foundation of Plato's epistemology, as we know it from his *Theaetetus*.

Saussure's ideas on language and linguistic structure have reached us through his *Cours de Linguistique Générale* (Saussure and Ferdinand 1916). The text was compiled by Charles Bally and Albert Sechehaye, after Saussure's sudden death in 1913, on the basis of notes they had collected from Saussure's lectures in Geneva. It was published for the first time in 1916.

As we have observed, the structuralist approach to the notion of meaning formulated by Saussure stands in sharp contrast with the Augustinian picture. However, the most explicit attack to the Augustinian picture and the view of meaning as reference was formulated, a few decades after Saussure's lectures, by Ludwig Wittgenstein in his *Philosophical Investigations* (Wittgenstein and Anscombe 1953). Wittgenstein's objections to the view of meaning as reference and, more precisely, his so-called private language argument will play an important role in part II of the book.

Another argument against the notion of meaning as reference was formulated by the philosopher Willard Van Orman Quine in the 1960s. In his book *Word and Object* (Quine 1960, especially ch. 2; see also ch. 2 of Quine 1969), Quine discusses the problem of *radical translation*. This is the problem faced by a translator (sometimes referred to as a radical translator) who attempts to translate a language that has never been translated or codified beforehand of which the interpreter has no knowledge. Quine demonstrates that the problem is, ultimately, an unsolvable problem. We exemplify his argument with a famous example of his. Imagine a radical translator observing for the first time a community of native speakers of an unknown language. Suppose the radical translator observes a native speaker of the language using the expression "gavagai" while pointing at a rabbit. Suppose, further, that the translator observes the same behavior repeated in other similar occasions. What can the translator conclude from these observations? The answer that first comes to mind is, of course, that "gavagai" means "rabbit". However, there are many other—in fact, potentially infinite—translations that are just as compatible with the actual use the native speaker is making of that word. "Gavagai" could be translated as a sentence—"Hey, there's a rabbit"—rather than a noun. It could mean "food" or "yummy". It could mean "There will be a storm tonight" because, unbeknownst to the translator, the community of native speakers believe in the superstition that rabbits bring storms. It could be used to refer to any of the constituting parts of the rabbit or to any of its temporal stages. It is easy to see that this exercise could go on for a long time—enough to prove that, ultimately, the translator will never be truly



able to conclude with certainty what the meaning of “gavagai” is. No matter how certain the interpreter is of a given interpretation, there is always another possible interpretation that is as compatible with the use that is made of that word by the native speakers as the one the interpreter believes to be correct. The problem of radical translation effectively exemplifies Quine’s general skepticism towards the notion of meaning and his preference for a behaviorist understanding of language, whereby the perceived meaningfulness of language is explained in terms of the actions its speakers perform and not in terms of the immaterial notion of meaning. We will not discuss Quine’s argument further in the book but the issue of indeterminacy of meaning will return prominently in part II of the book, particularly in relation to Wittgenstein’s argument against private language.

The example “the gostak distims the dosh” is modeled over an example originally presented by Charles Kay Ogden and Ivor Armstrong Richards in their book *The Meaning of Meaning* (Ogden and Richards 1923). The fact that part of the meaning of a complex linguistic expression can be recovered from its bare grammatical structure has been also exploited for literary purposes. An extraordinary example is Lewis Carroll’s “Jabberwocky”, a non-sense poem included in his 1871 book *Through the Looking Glass, and What Alice Found There*. This is the first stanza of the poem:

Twas brillig, and the slithy toves  
 Did gyre and gimble in the wabe:  
 All mimsy were the borogoves,  
 And the mome raths outgrabe.

The poem is made of non-sense words but its grammar is that of proper English. In the end, it is impossible to interpret the poem with precision, because its content words are non-sense words. Yet, its grammatical structure is enough to suggest the poem is about the killing of a creature—“the Jabberwock”—and, as many have suggested, the conflict between good and evil.

## References

- Ogden, C. K., & Richards, I. A. (1923). *The meaning of meaning*. London: Kegan Paul.
- Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.
- Quine, W. V. O. (1969). *Ontological relativity and other essays*. New York: Columbia University Press.
- Saussure, D., & Ferdinand. (1916). *Cours de linguistique générale, publié par Ch. Bally et A. Sechehaye avec la collaboration de A. Riedlinger*. Paris: Payot.
- Wittgenstein, L., & Anscombe, G. E. M. (1953). *Philosophical investigations*. London: Basic Blackwell.

## Chapter 4

# Meaning and Compositionality



We owe to Frege the idea that meaning is *compositional*. However we decide to characterize the notion of meaning, it must have combinatorial properties. The meanings of individual words must be such that they can combine together in richer and richer grammatical constructions to produce richer and richer meanings. This insight is known as Frege's *principle of compositionality* and is formulated as follows:

### Principle of Compositionality

The meaning of a complex linguistic expression E is a function of:

1. The meaning of the simple linguistic expressions occurring in E
2. The grammatical structure of E.

The principle tells us that the meaning of a complex expression—any grammatical combination of two or more words—is the product of two ingredients: the individual meanings of the individual words occurring in it and the grammatical structure that holds these words together. In Frege's view, grammatical relations are the glue that holds together the meaning of different words as they occur in complex grammatical structures.

As Frege's principle of compositionality will be our first step towards constructing an actual theory of linguistic meaning, it is useful to look at it in some details. First, let us consider the terminology it employs. The principle refers to those elements of language that express meaning as *linguistic expressions* and distinguishes between two types of such expressions, *simple* and *complex*. Simple expressions are the smallest units in a language that express meaning—those expressions that cannot be broken down into even smaller pieces without losing the capacity to express meaning. Linguists generally refer to these units as *morphemes*. In this book, we

will simplify things and identify simple expressions with individual *words*. The simple expressions of English are, henceforth, expressions such as “table”, “Kazimir”, “admire”, “tall”, “the”, “and”, “well”, etc. Complex expressions, in contrast, are grammatical linguistic expressions that result from the combination of two or more simple expressions. “Kazimir admires Frida”, “Adam and Eve”, and “read the book” are all complex expressions.

Both simple and complex expressions bear meaning. However, they do so in different ways. The association between a simple expression and its meaning is the result of a *stipulation* and, therefore, *arbitrary*. There is no inherent property of the sequence of sounds corresponding to a simple word—say “table”—that grants us knowledge of its meaning. As speakers of English, we simply have to learn what “table” means, either by receiving explicit instruction, or by inferring its meaning from how the word is used by other speakers. Of course, there are historical reasons why at some point English speakers converged on using that particular sequence of sounds to refer to tables. However, knowledge of such reasons does not make us any more proficient in using and understanding the word “table”. All we really need to know, if we want to be able to use and understand the word “table” in English, is the association between the word and what it means. This is not the case for complex expressions. Consider, as an example, the following sentence:

(1) Kazimir admires Frida

It may very well be the case that you never encountered this complex expression before in your experience of the English language. Nor are you likely to have received explicit instructions about the meaning of this specific sentence. Yet, this does not prevent you from understanding what the expression means: A certain person called Kazimir admires a certain other person called Frida. The reason why you can grasp the meaning of (1), even though you have never encountered it before, is that you know how to combine together the individual meanings of the individual words “Kazimir”, “admires”, and “Frida” to obtain the meaning of (1). The association that (1) has with its meaning is, therefore, not the result of a mere act of stipulation. Rather, it is the outcome of a *procedure*, which allows proficient users of English to combine together the meanings of the simple expressions in (1) to obtain the meaning of the whole complex expression.

Frege’s principle of compositionality is important because it represents the first attempt at offering a formal description of the procedure that delivers the meaning of complex expressions from the meaning of simple ones. The principle, in fact, is meant to capture a fundamental property of natural language: its *creativity*. Language is a system that allows speakers to rely on a finite amount of information—information about the meaning of the simple expressions in their language—to generate a potentially infinite amount of novel complex expressions and, correspondingly, an infinite amount of complex meanings. It is, to use a different term, a *generative* system—a system that exploits a finite amount of resources to generate a potentially infinite range of outcomes. More precisely, we can characterize a natural language as the product of two main components: (a) a vocabulary or, as we will call it from now on, a *lexicon*—a storage of simple expressions and their meanings; and (b) a *grammar*—a procedure that allows language users to combine the meaning of

simple expressions into complex ones. These two components alone provide speakers with everything they need to use language creatively. We can imagine language as a game of bricks. We have a box of bricks and a procedure to combine them together. With those, we can construct anything we want.

## References and Remarks

The principle of compositionality is commonly attributed to Frege although, in fact, he never explicitly formulated it. The principle is rather inferred from different passages in his writings on the topics of language and meaning (Frege 1891, 1892, 1923). To this we should add that it is questionable whether Frege himself would have endorsed the principle of compositionality in the form in which it is commonly endorsed today. In one of his earlier writings (Frege 1884), Frege mentions another principle, which came to be known as the principle of contextuality, whereby the constituents of complex linguistic expressions do not have meaning in isolation but only as a function of the grammatical context in which they occur. Such principle is in sharp contrast with the principle of compositionality as we have presented it in this chapter and, in fact, more akin to the structuralist view of meaning that we discussed in the previous chapter. We refer the readers interested in this issue to Pelletier’s article “Did Frege Believe Frege’s Principle?” (Pelletier 2001).

A number of eminent thinkers have endorsed the idea that natural language is compositional. Amongst them, we should quote Jerry Fodor’s claim that compositionality is a “not-negotiable” primitive of human language (Fodor 2001). Others, however, have contested the idea that the natural language is compositional. For an overview of the challenges to the notion of compositionality and potential solutions, we recommend Zimmerman’s article “Compositionality Problems and How to Solve Them” (Zimmermann 2012). Worth mentioning, in this context, is the issue of idioms. Idioms are complex expressions whose meaning is not the straightforward product of their parts. Consider, as an example, the idiomatic expression “kick the bucket”. In its most typical use, the expression does not express the actual kicking of an actual bucket. Its literal meaning is rather overwritten by a conventionally established meaning that is established lexically rather than constructed compositionally. The problem is compounded by the fact that there are idioms that are partially compositional—for example, “spill the beans”. We refer the reader interested in this issue to Jaume Mateu and María Teresa Espinal’s article “Argument structure and compositionality in idiomatic constructions” (Mateu and Teresa Espinal 2007).

## References

- Fodor, J. A. (2001). Language, thought, and compositionality. *Mind and Language*, 16, 1–15.  
 Frege, G. (1884). *Die Grundlagen der Arithmetik. Eine logisch mathematische Untersuchung über den Begriff der Zahl*. Breslau: Verlag von Wilhelm Koebner.

- Frege, G. (1891). *Funktion und Begriff. Vortrag, gehalten in der Sitzung vom 9. Januar 1891 der Jenaischen Gesellschaft für Medizin und Naturwissenschaft*. Jena: Hermann Pohle.
- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und Philosophische Kritik*, 100(1), 25–50.
- Frege, G. (1923). Logische Untersuchungen. Dritter Teil: Gedankengefiige. *Beitrdge zur Philosophic des Deulschen Idealismus*, 3, 36–51.
- Mateu, J., & Teresa Espinal, M. (2007). Argument structure and compositionality in idiomatic constructions. *The Linguistic Review*, 24, 33–59.
- Pelletier, F. J. (2001). Did Frege believe Frege's principle? *Journal of Logic, Language and Information*, 10(1), 87–114.
- Zimmermann, T. E. (2012). Compositionality problems and how to solve them. In *Oxford handbook of compositionality* (pp. 81–106). Oxford: Oxford University Press.

## Chapter 5

# Meaning and Grammar



Frege's principle of compositionality tells us that grammatical structure is the glue that holds together simple meanings into complex ones. We can appreciate the importance of Frege's principle by comparing the sentences (1) and (2).

- (1) Kazimir admires Frida
- (2) Frida admires Kazimir

The two sentences are made of the same simple expressions, the words “Kazimir”, “Frida”, and “admires”. Yet, they convey different meanings. Sentence (1) tells us that Kazimir has admiration for Frida; sentence (2) that Frida has admiration for Kazimir. What determines that (1) and (2) express different meanings despite the fact that they are constructed from the very same ingredients? Quite obviously, what determines the difference between (1) and (2) is *how* the same simple expressions are combined together in the two sentences—that is, to use the terminology of Frege's principle, their *mode of grammatical combination*. In the case of (1), “Kazimir” is the grammatical subject of the verb “admires” whereas “Frida” is its direct object. In the case of (2), “Frida” is the subject whereas “Kazimir” is the object.

What is then the mode of grammatical combination that holds together simple expressions into complex ones in natural language? What are the procedures that allow natural language speakers to combine words into sentences? *Syntax* is the branch of linguistics that investigates just that.

The first thing we must understand to appreciate the work of syntacticians is the very notion of *grammar*. Ordinarily, we understand a grammar to be the set of conventions that determine what is the proper use of a language. The notion of grammar syntacticians have in mind when studying natural language syntax has a more technical flavor. Consider the following three simple expressions of English: “I”, “him”,

and “admire”. There is only one complex expression of English that can be constructed from them:

(3) I admire him

Sentence (3) is a grammatical sentence of English. Any other combination of these three simple expressions is ungrammatical: “Him admire I”, “admire him I”, “I him admire”, etc. are all unacceptable combinations in English. The grammar of English, as intended by syntacticians, is the set of rules of combination of simple expressions that allows speakers of English to construct all grammatical complex expressions of English but none of the ungrammatical ones. It is the set of rules that decides that (3) is a possible expression of English but any other combination of the same words is not. Through their work syntacticians aim at identifying and describing, as explicitly as possible, the grammars of natural languages and, more generally, the grammatical principles at the basis of all natural languages.

## Finite-State Grammar

There are different ways to express the grammatical rules of a language. A simple way consists in attaching constraints to the simple expressions in the lexicon of the language, restricting the combinatorial possibilities of these simple expressions to those that are actually attested in the language. We can visualize a grammar of this type as a puzzle. Each simple expression is a piece characterized by a specific shape. Because of their specific shape, simple expressions can combine in some ways but not others. As an example, imagine that the three simple expressions “I”, “him”, and “admire” correspond to the three shapes in Fig. 5.1. There is only one way to combine these three pieces together, which is the way exemplified in Fig. 5.2. Any other combination is impossible, because the shape of the pieces does not allow it.



**Fig. 5.1** The three English words “I”, “him”, and “admire” represented as pieces of a puzzle with different shapes



**Fig. 5.2** The only possible combination of the words “I”, “him”, and “admire” that can be obtained on the basis their shape



**Fig. 5.3** The three English words “Kazimir”, “paints”, and “paint” represented as pieces of a puzzle with different shapes

Syntacticians refer to this type of grammar as a *Finite-State Grammar*. The name reflects the fact that, when we are constructing a sequence of simple expressions in such a grammar, we can always tell after each step in the procedure—that is, after each state—what expressions can follow and what expressions cannot. After we have put our first piece of the puzzle on the table, we can tell which pieces can come next and which ones cannot. To go back to our example, if the piece (state) that we are currently considering is the simple expression “admire”, the grammar tells us that the next piece can be “him”, but not “I”.

### Phrase-Structure Grammar

A finite-state grammar of the sort just described offers a simple and effective way to appreciate the combinatorial side of the notion of grammar. However, it is questionable whether a finite-state grammar is sufficient to describe the grammar of an actual natural language, such as English. This point was made for the first time by the linguist Noam Chomsky in the mid 1950s. His observations would become fundamental in shaping the field of natural language syntax in the decades to come.

As we saw, we can legitimately talk about a finite-state grammar if, and only if, at any state in the step-by-step construction of a complex expression, we can predict which simple expressions can realize the following state. Suppose we begin the construction of a new complex expression with the simple expression “Kazimir”. “Kazimir” realizes our first state. How can we proceed? Consider two potential candidates, “paints” and “paint”. Our knowledge of speakers of English tells us that “paints” is a possible continuation but “paint” is not:

- (4a) Kazimir paints
- (4b) \*Kazimir paint

English speakers judge (4a) as a grammatical sentence of English, but (4b) as an ungrammatical one. As is common practice in syntax, we mark the ungrammatical status of (4b) by putting a star in front of it. We can capture the contrast between (4a) and (4b) in a finite-state grammar by constructing a rule that allows us to use “paints”, but not “paint” as a possible continuation for “Kazimir”. In our puzzle grammar we can capture this rule by constructing the pieces as in Fig. 5.3. The pieces are designed in such a way that “Kazimir” and “paints” can combine in a sequence, but “Kazimir” and “paint” cannot.



So far so good. However, consider now the following pair of sentences:

- (5a) Few friends of Kazimir paint  
 (5b) \*Few friends of Kazimir paints

Speakers of English judge sentence (5a) as a grammatical sentence, but (5b) as an ungrammatical one, as indicated by the star at the beginning of the sentence. This is a problem for the rule we have designed because the rule predicts the opposite pattern. According to our finite-state rule, “Kazimir” can be followed by “paints” but not by “paint”. What we find in (5) is the opposite: “Kazimir” can be followed by “paint”, as demonstrated by the grammaticality of (5a), but it cannot be followed by “paints”, as demonstrated by the ungrammaticality of (5b).

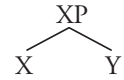
What went wrong? To address the problem, we must look at it from a different angle. The reason why “Kazimir” must be followed by “paints” in (4) is fairly obvious: “Kazimir” is the subject of the sentence and it is a third-person singular noun. In English, a verb in the present tense must agree with a third-person singular subject by displaying a final *-s*. Similarly, the reason why “paint”, but not “paints”, is grammatical in (5) is that the subject of the sentence is not the simple expression “Kazimir”, which is singular, but the complex expression “few friends of Kazimir”, which is plural. The main difference between (4) and (5) is, therefore, that in (4) the subject of the sentence is a simple expression, “Kazimir”, whereas in (5) it is a complex one, “few friends of Kazimir”.

This explanation seems on the right track. The problem is that it cannot be translated into a rule of a finite-state grammar. A finite-state grammar, as we saw a moment ago, proceeds in a step-by-step fashion: The choice of every new word in a sequence is determined by the word that immediately precedes it. If English was governed by a finite-state grammar, we should be able to explain the grammaticality of “paint” in (5) solely on the basis of the word that immediately precedes it. This, however, is not what we find. The pair of sentences in (5) shows us that, in order to explain the grammaticality of “paint”, we must consider not just the word immediately preceding it, but the whole complex expression “few friends of Kazimir”.

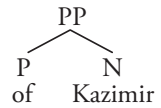
In order to fully express the grammatical potential of a natural language such as English, we need something more powerful than a finite-state grammar. A type of grammar that can handle the problem raised by (5) is a *phrase-structure grammar*. Whereas a finite-state grammar proceeds by chaining individual words one after the other in a sequential order, a phrase-structure grammar proceeds by combining pairs of expressions—no matter if simple or complex—into units of larger and larger complexity. These units are typically referred to as *phrases*, hence the name phrase-structure grammar. In the technical sense intended by syntacticians, a phrase is a cluster of words that together perform a grammatical function. The complex expression “few friends of Kazimir”, to return to our example sentence in (5a), is a phrase because it is a group of simple expressions that together perform the grammatical function of subject of the sentence.

At the heart of a phrase-structure grammar there is a basic combinatorial operation which takes two expressions—no matter if simple or complex—and delivers as a result a new complex expression that combines the two original expressions into a

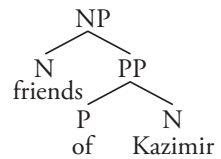
**Fig. 5.4** The Merge operation takes two expressions X and Y—be them simple or complex—and combines them into a phrase XP



**Fig. 5.5** The preposition (P) “of” merges with the noun (N) “Kazimir” in the Prepositional Phrase (PP) “of Kazimir”



**Fig. 5.6** “Friends” merges with “of Kazimir” in the noun phrase (NP) “friends of Kazimir”



more complex syntactic unit. In syntax, it is customary to refer to this combinatorial operation as *Merge*. We can represent the Merge operation as in Fig. 5.4. The figure must be read from the bottom up. Given two expressions X and Y—simple or complex—Merge combines them into a larger phrase, called XP. Merge is, therefore, the mechanism through which two existing grammatical objects can be combined into a new, more complex one.

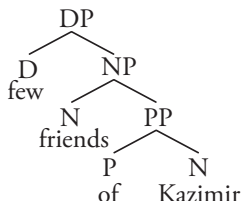
As the figure suggests, Merge is an asymmetric operation. Whenever two grammatical expressions are combined by Merge, one of them takes priority over the other in determining the grammatical status of the resulting phrase. The figure above tells us that X is the most important of the two elements because it is the one that decides the grammatical status of the resulting phrase and, accordingly, provides a label (XP) to the resulting phrase. This element is called the *head* of the phrase and it is the element that imposes combinatorial constraints on the other. It is X that determines whether Y is an admissible object to combine with.

The best way to appreciate the functioning of a phrase-structure grammar is to see it in action. Let us consider in some details how the Merge operation proceeds to combining the individual simple expressions “few”, “friends”, “of”, “Kazimir”, and “paint” into phrases of larger and larger complexity up to the point where they contribute the sentence “few friends of Kazimir paint”. Let us consider this process in some more detail by examining each of its derivational steps.

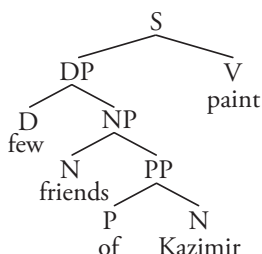
The first step, from the bottom up, consists in combining “of” with “Kazimir”. This is exemplified in Fig. 5.5. “Of” is a preposition and this is symbolized in the figure below by a P (for preposition) immediately above it. “Kazimir” is a noun, symbolized below as N. The Merge operation combines these two simple expressions into a phrase. The head of the resulting phrase is the preposition “of” because it is the preposition that imposes a combinatorial constraint—“of” demands to be combined with a noun. Accordingly, the resulting phrase is a prepositional phrase (symbolized in the figure as PP).

The second step consists in combining the PP “of Kazimir” with the noun “friends” to form a Noun Phrase (NP), as exemplified in Fig. 5.6. In this case, the

**Fig. 5.7** “Few” merges with “friends of Kazimir” in the determiner phrase (DP) “few friends of Kazimir”



**Fig. 5.8** “Few friends of Kazimir” merges with the verb (V) “paint” in the Sentence (S) “few friends of Kazimir paint”



head is the noun “friends” because it is “friends” that licenses the combination with a prepositional phrase.

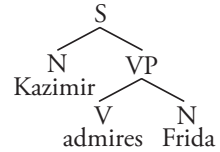
The third step consists in combining the noun phrase “friends of Kazimir” with the determiner “few” to produce a Determiner Phrase (DP), as exemplified in Fig. 5.7. In this case, the head is the determiner “few”, which is the element selecting the Noun Phrase.

In the last step, the DP “few friends of Kazimir” is combined with the verb “paint” to contribute the whole sentence, as exemplified in Fig. 5.8. We will adopt the view that sentences are somewhat special phrases that are always the combination of a nominal element—the subject—and a verbal element—the predicate. Accordingly, we will refer to those phrases that are also sentences simply as S. Today, syntacticians use more sophisticated analyses of the notion of sentence. For our purposes, however, there is no need to make things more complicated than necessary.

We see from this exercise how a phrase-structure grammar works by combining pairs of linguistic expressions, complex and simple, into larger phrases. Within this type of grammar, we can easily formulate a rule that allows us to explain why “few friends of Kazimir” combines with “paint” but “Kazimir” combines with “paints”. All we need to do is to formulate a restriction on how a nominal expression, no matter if complex or simple, can combine with a present tense verb to provide a sentence. If the nominal expression is third-person singular—such as in the case of “Kazimir”, the verb is “paints”. If the expression is not third person singular—such as in the case of “few friends of Kazimir”—the verb is “paint”.

We can also appreciate another important property of the Merge operation. It is a recursive type of operation. Once we have merged two expressions together, we can use the resulting expression as the input of a new merge operation. This is an

**Fig. 5.9** Phrase-structure of the sentence “Kazimir admires Frida”



important property because it allows us to capture the generative power of natural language grammars—the fact, that is, that natural language grammar allows us to produce an infinite amount of complex expressions on the basis of a finite number of simple expressions and rules of combination.

The most important conclusion, for the purposes of our book, is that natural languages have a degree of grammatical complexity which requires a phrase-structure grammar. A procedure that chains words one after the other in a simple linear fashion, such as a finite-state grammar, is not enough. What we need in order to describe the mode of combination of simple expressions into complex ones is a procedure that combines words into larger and larger phrases—that is, a phrase-structure grammar.

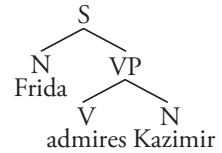
## Grammar and Compositionality

The lesson we draw from this brief excursion into natural language syntax is that the mode of grammatical combination exploited by natural language follows a specific pattern, that of a phrase-structure grammar. This observation is important because, according to Frege’s principle of compositionality, this is also the way in which the meanings of the simple expressions combine together to provide the meaning of complex expressions. Consider again the sentences in (1) and (2), repeated below.

- (1) Kazimir admires Frida
- (2) Frida admires Kazimir

As we saw, these two sentences express different meanings. Frege’s principle of compositionality tells us that these meanings are the product of two factors: (a) the meanings of the individual words occurring in the two sentences and (b) their mode of grammatical combination. As we observed, since the two sentences are made of exactly the same simple expressions, their difference in meaning must be the product of the different modes in which these simple expressions are combined grammatically within the two sentences. Thanks to what we have learned in this chapter, we can now characterize the difference between the grammatical structures of the two sentences in precise terms. Sentence (1), on the one hand, is constructed by first merging the verb “admires” with its direct object “Frida”, and then by merging the resulting Verb Phrase (VP) “admires Frida” with the subject “Kazimir”, as exemplified in Fig. 5.9. Sentence (2), on the other hand, is constructed by first merging the verb “admires” with the direct object “Kazimir” and then by merging the resulting

**Fig. 5.10** Phrase-structure of the sentence “Frida admires Kazimir”



Verb Phrase “admires Kazimir” with the subject “Frida”, as exemplified in Fig. 5.10. The two sentences are produced by combining the same words in different orders. It is this difference that determines their difference in meaning.

Frege’s principle of compositionality may not tell us yet what meaning is but suggests two fundamental properties that meaning must meet if it is to be made compatible with how natural language grammar works. The first is that meaning originates with individual words. The second is that it is compositional. The individual meanings of individual words must be such that they can combine with one another, when occurring in larger grammatical structures, by following the specific order imposed by the phrase-structure grammar that holds them together. Our next challenge will be, henceforth, that of characterizing the notion of meaning so that it can meet these two properties. As we will see in the next chapter, in order to reconcile the observation that words have distinctive meanings with the observation that meaning is compositional, we must look at meaning as a *formal* object. As Malevich reverted his attention from what images are about to their formal constituting elements, so we must do for language. This, then, will be our next question: What is the form of meaning?

## References and Remarks

Chomsky’s article “Three models for the description of language”, published in 1956 (Chomsky 1956), represents the first attempt at providing a formal description of the computational properties of a natural language grammar. In the article (see also Chomsky 1957, 1959; Chomsky and Schützenberger 1963), Chomsky provides formal definitions of different types of grammar, characterized by different degrees of complexity, and identifies phrase structure grammars as the type needed for capturing the complexity of natural language structure. The organization of the different types of grammar on the basis of their formal properties and computational complexity came to be known as *Chomsky hierarchy* and has been highly influential, beyond the domain of linguistics, in the fields of computer science and artificial intelligence.

Chomsky’s earlier work set the foundations of the field of modern generative grammar and the development of different theories of natural language structure, such as the so-called Standard Theory (Chomsky 1965) and the theory of Principles and Parameters (Chomsky 1981). To be sure, as observed by Chomsky himself as early as in his 1956 article, there is more to natural language structure than just

phrase-structure, as some structures are the products of operations of *transformation*. These are operations that manipulate existing phrase structures to derive new ones by displacing their constituents. An example of such operation is *wh*-movement. This is the operation responsible for the displacement of *wh*-phrases at the front of *wh*-questions in languages such as English. In a more recent incarnation of Chomsky's framework, known as the Minimalist Program (Chomsky 1995), an attempt is made at eliminating transformations by distinguishing between two types of merge operations: *external* merge, which is merge between an existing phrase structure XP and an item Y taken from the lexicon, and *internal* merge, which is merge between an existing phrase structure XP and an item Y that is a constituent element of XP itself.

## References

- Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, 2(3), 113–124.
- Chomsky, N. (1957). *Syntactic structure*. The Hague: Mouton.
- Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control*, 2(2), 137–167.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. (1995). *The minimalist program*. Cambridge: MIT Press.
- Chomsky, N., & Schützenberger, M. P. (1963). The algebraic theory of context-free languages. *Studies in Logic and the Foundations of Mathematics*, 35, 118–161.

## Chapter 6

# Models and Formal Languages



Frege was the first of a number of thinkers who, across the end of the nineteenth and the beginning of the twentieth centuries, expressed discontent towards the use of natural languages as the means to formulate philosophical and scientific ideas. Frege reasoned as follow. Natural languages are easily prone to ambiguities and can lead to errors and misinterpretations. Yet, they are the main instrument of analysis and argumentation in a number of disciplines, including, crucially, philosophy. It is not surprising, then, that some of the most fundamental questions in philosophy remain unanswered, despite the effort of many brilliant scholars across the centuries. Perhaps, what has prevented progress is not that the issues are unsolvable but that they were formulated in a language that is not sufficiently rigorous and precise.

Frege's solution to this problem was not dissimilar in spirit to that adopted by Malevich and the Zaum poets to address the problem of creativity in art. Frege's inspiration came from the disciplines of arithmetic and geometry, which he had studied at great length in his early years. Key to the progresses achieved in these disciplines, Frege observed, was the adoption of the language of mathematics, which is rigorous and precise. Frege decided then to explore the core workings of that language with the aim of identifying its fundamental formal structure and eventually providing a formula to construct a language that has the rigor of mathematics but, at the same time, the generality required to address any subject matter, from algebra to philosophy. By doing so, Frege initiated the study of *formal languages*. His ideas were developed in the decades to follow by a number of mathematicians, logicians, and philosophers, such as Bertrand Russell, Ludwig Wittgenstein, and Alfred Tarski. Several decades later, the same ideas contributed to the formulation of the first formal theory of natural language meaning.

In this chapter, we will familiarize ourselves with the notions of formal language and model of interpretation. In the chapters that will follow, we will see how these notions provide the basis for understanding natural language meaning and compositionality.

## Models

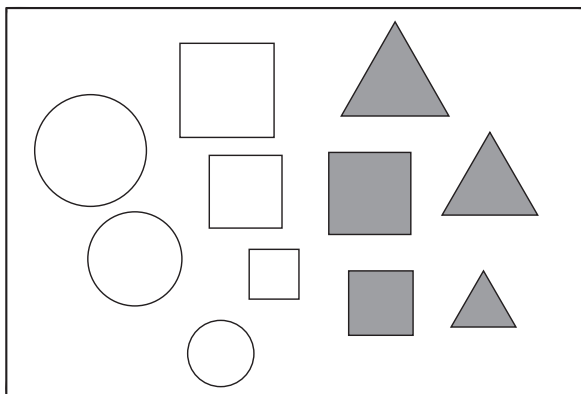
A formal language is, first and foremost, formal. It is a language whose lexicon, grammatical rules, and rules of interpretation are defined from the start in the most explicit and rigorous way. It is also an artificial language, whose rules are established from the very start by its creator. Setting up a simple formal language and a corresponding model of interpretation is easy. Doing so will allow us to explore the mathematical mechanisms at its core. We will begin, in this section, by introducing the notion of *model*.

In a number of sciences, natural and theoretical alike, understanding a phenomenon corresponds to constructing a model of it. A model is, in its most basic form, a collection of objects organized on the basis of their properties. To achieve a more concrete grasp of this notion, let us consider a simple example. Imagine for a moment that we were interested in studying the properties of a small group of geometric figures. We can begin constructing a model by organizing these figures in a set, also called the *domain* of the model. For the sake of the example, let us assume that the domain of our model corresponds to the collection of geometric figures in Fig. 6.1.

The figures in this model are characterized by two main properties: color (gray and white), and shape (triangle, square, and circle). In Fig. 6.2 we have organized the objects in the domain by grouping them into smaller sets, on the basis of these properties. We have grouped them on the basis of their color, gray and white (these are the sets contained within the gray areas in the figure). We have also grouped them on the basis of their shape, that is, into circles, squares, and triangles (these are the sets contained within continuous lines).

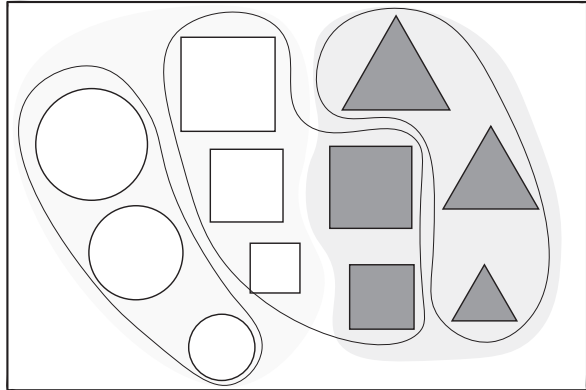
This simple organization of objects into sets is already a powerful mathematical tool. In our simple model, we can already observe how organizing objects into sets allows us to discover the logical relations that hold amongst them. We can observe, for example, that the set of circles is included within the set of white objects. We will say that the set of circles is a *subset* of the set of white objects. Similarly, we

**Fig. 6.1** The domain of a model consisting of geometric figures characterized by two different colors (gray and white) and three different shapes (triangle, square, and circle)





**Fig. 6.2** The geometric figures in the domain of the model are organized on the basis of their color, gray and white (these are the sets contained within the gray areas in the figure) and their shape, circle, square, and triangle (these are the sets contained within continuous lines)



can observe that the set of triangles is a subset of the set of grey objects. These relations between sets bring about corresponding logical relations. We can say that, in the model we are considering, the property of being a circle logically *entails* that of being white—if an object is a circle, then it is also white—and the property of being a triangle logically *entails* that of being grey—if an object is a triangle, then it is also grey.

The strategy of organizing objects into sets is so simple and abstract that it can be used as a general analytic tool for the study of all sorts of things. At the core of a model, there are two primitive notions: the notion of an *object* and the notion of a *set*. Both notions are very abstract and, therefore, very general. In fact, anything can be an object in a model as long as it can be conceived as a *mathematical quantity*, that is, something that can belong to a set. Model-theoretic objects can be human beings, neurons, paintings, prices, money, dreams, beliefs, chairs, telephone calls, molecules, kisses, passengers, bottles of wine, music festivals, internet users, books read, books written, supermarket products, relationships, bits of information, you name it.

## The Language of Models

Now that we have a model, we can use it for a further exercise: setting up a formal language for the purpose of describing it. This exercise will be useful when, in the coming chapters, we will attempt at providing a formal theory of natural language meaning.

Minimally, a formal language is a rigorously organized system of symbols that is capable of telling us which objects belong to which sets in a model. The first step we must take when constructing a formal language consists in providing a vocabulary or, as we will say from now on, a *lexicon*. A lexicon is a set of symbols that we use to refer to the objects and sets in the model. The lexicon we will use in our

example includes two types of simple expressions, lower-case letters and upper-case letters:

*Lexicon*

- a. Lower-case letters: *a, b, c, d, e, f, g, h, i, j, and k*;
- b. Upper-case letters: *C, S, T, G, and W*.

The second step in setting up our formal language requires providing a *grammar*, which specifies how the symbols in the lexicon can combine into complex expressions. For the time being, we will adopt the following rules:

*Grammar*

- a. Any lower-case letter followed by any upper-case letter is a grammatical combination;
- b. Any other combination is ungrammatical.

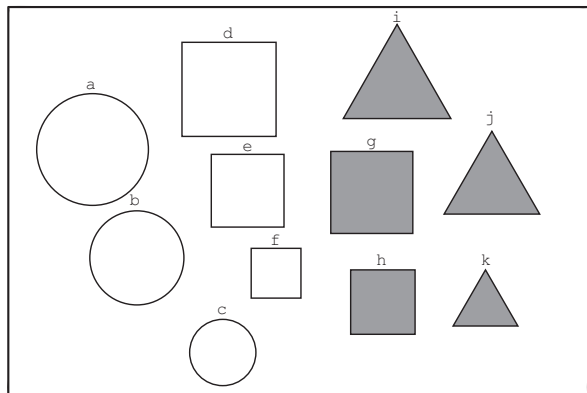
These grammatical rules determine which combinations of the symbols in the lexicon are allowed and which are not. The rules allow us only to construct sequences of a lower-case letter followed by an upper-case letter, such as *bS*, *kT*, and *aW*. Any other type of combination, such as *dCf*, *Gc*, or *TSj*, is excluded.

The third and final step requires providing a procedure for interpreting the expressions produced by the formal language. First, we must assign meanings to the individual symbols. We will use the lower-case letters to refer to the individual objects in the model, as indicated in Fig. 6.3.

We will use the upper-case letters to refer to the sets in the model. We will use *W* to refer to the set of white objects, *G* to the set of grey objects, *C* to the set of circles, *S* to the set of squares, and *T* to the set of triangles.

The final step in setting up our formal language consists in providing a procedure for assigning a meaning to the combinations of symbols that are allowed by the grammar of the language. To this end, we will adopt the following rules of interpretation:

**Fig. 6.3** The figure indicates which lower-case letters in the lexicon of the formal language refer to which figures

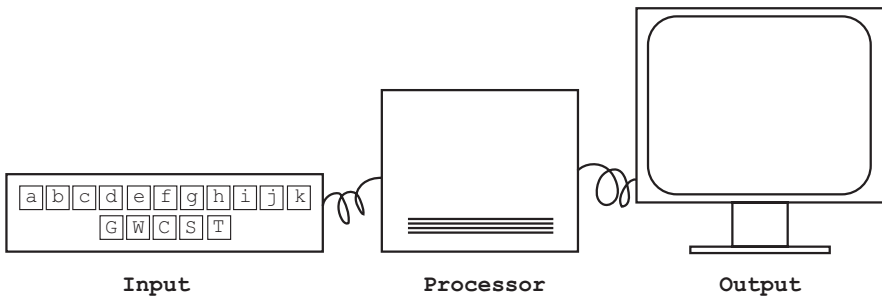


*Interpretation*

- a. A complex expression of the form *lower-case letter + upper-case letter* is assigned the truth-value 1 if, and only if, the object referred to by the lower-case letter belongs to the set referred to by the upper-case letter.
- b. A complex expression of the form *lower-case letter + upper-case letter* assigned the truth-value 0 if, and only if, the object referred to by the lower-case letter does not belong to the set referred to by the upper-case letter.

According to these rules, the complex expressions of our language refer to a number—0 or 1—depending on whether the object they refer to by means of the lower-case letter belongs or does not belong to the set designated by the upper-case letter. We will refer to this number as a *truth-value*. Consider, as an example, the grammatical expression *aG*. According to our rule of interpretation, the expression refers to the truth-value 1—and therefore is true—if, and only if, the object referred to by *a* belongs to the set referred to by *G* but refers to the truth-value 0—and, therefore, is false—if, and only if, object *a* does not belong to set *G*. As the object referred to by *a* does not belong to the set *G* of grey things, the expression is false. Put more informally, the expression corresponds to the claim that *a* is grey. Hence, it is true in the model it describes if, and only if, *a* belongs to the set of grey objects. Otherwise, it is false.

We can think of the formal language we have just designed as a sort of machine comprising three components, as exemplified in Fig. 6.4: (a) an *input* device (such as a keyboard); (b) a *processor*; and (c) an *output* device (such as a screen). The input device allows us to feed input information to the machine, the processor analyzes the input, and the output device delivers the result of the analysis. Whenever we type a grammatical expression in the machine’s keyboard, the processor checks whether that expression is true or false in the model. If we type the expression *aG*, the machine checks whether the object referred to by *a* belongs to the set referred to by *G*. If the machine finds that *a* does belong to the set of grey objects, it delivers the number 1 as its output. In this way, we know that the expression is true. If the machine finds that *a* does not belong to the set of grey objects, it delivers the number 0. In this way, we know that the expression is false. Since, in the model we are



**Fig. 6.4** A machine consisting of (from left to right) an input device, a processor, and an output device. Such a machine could implement the formal language we have designed in this chapter

considering, figure  $a$  does not belong to set  $G$ , the machine will display the number 0 on the screen, telling us that the expression we typed is false:  $a$  is not grey.

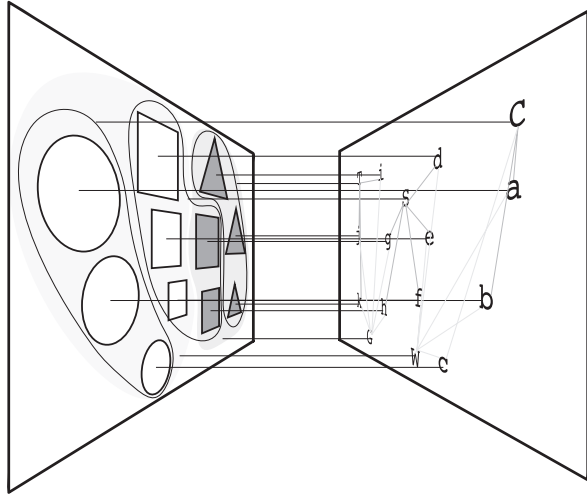
## Language as a Mirror

The study of formal languages initiated by Frege had powerful consequences in a number of domains, from mathematics to philosophy. It is, indeed, at the foundation of the modern notion of a machine and of computer science as we know it today. It has also been, as we will see in the coming chapters, of fundamental importance for the development of modern theories of natural language meaning. For the time being, it is useful to briefly reflect on the notion of meaning that emerges from the study of formal languages. When constructing our language, we provided its interpretation by defining, for each expression in the language, simple or complex, a corresponding reference in the model of interpretation—an object, a set, or a truth-value. Logicians typically refer to the model against which a formal language is interpreted as its *world* (or, sometimes, *universe*) of interpretation. A formal language that is constructed on the basis of such notion is, mathematically speaking, a *homomorphism*. Imagine yourself standing in front of a mirror (assume the mirror is faithful in providing its reflection and not a tricked mirror). On the other side of the glass, there is a reflection of yourself. Whenever you move one of your arms, the reflection does it too. Whenever you turn your head to one side, the reflection does it too. For every property of your figure that faces the surface of the mirror, there is a corresponding property of the reflection. And for every feature of the reflection, there is a corresponding feature of your figure. What the mirror does is to create a homomorphism between two different classes of objects: your body, on one side, and the distribution of light on the mirror, on the other. The two objects, although different in their constitution, behave in perfect harmony: Any variation in one of them finds a corresponding variation in the other.

The function performed by a formal language is similar to that performed by the mirror. It describes its world of interpretation by producing a “reflection” of it. For each object and set in the model, we define a corresponding simple expression in the language. For each relation in the model, we define a corresponding complex expression in the language. The language, hence, mirrors the model. We can visualize the overall structure of a formal language and its world of interpretation as in Fig. 6.5. On the left side, we find the model—a set of objects and relations among them. On the right side, we find the formal language—a set of symbols and grammatical relations among them. The connecting lines tell us which symbols of the language refer to which objects in the model. Once again, we see how the view of language as a mean of reference to a domain of objects is inextricably intertwined with the notion of truth. How good a language is in performing its function depends on how truthfully it reflects its world of interpretation.

At the beginning of the 1970s, an ingenious mathematician and logician performed a theoretical experiment that would be seminal in shaping modern natural

**Fig. 6.5** A representation of the system of interpretation of a formal language as a homomorphism. The left side represents the model against which the language is interpreted—a set of objects and relations among them. The right side represents the formal language—a set of symbols and grammatical relations among them. The connecting lines indicate the reference of the symbols in the model



language semantics. He wondered whether the view of meaning that is used in the domain of formal languages could be extended also to natural languages. Not only he found that the overall mathematical structure that governs formal languages can be extended to natural ones but also that, by doing so, he could provide a straightforward account of their compositional nature. This logician was called Richard Montague and it is to his work that we turn in the following chapters.

## References and Remarks

The modern notion of formal language was first introduced by Frege in his *Begriffsschrift* (Frege 1879) with the specific purpose of studying the logical foundations of mathematics. Frege's work on formal languages would be pivotal to the development of modern analytic philosophy and to the work of thinkers such as Bertrand Russell and Ludwig Wittgenstein—in particular, the Wittgenstein's of the *Tractatus Logico-Philosophicus* (Wittgenstein 1921).

The formal language we have considered in this chapter is a simplified form of Predicate Logic that decomposes expressions into subjects, typically represented by lower-case letter, and predicates, represented by upper-case letters. It is common to construct formulas of predicate calculus according to the order upper-case letter, lower-case letter (for example,  $Ga$ ,  $Wc$ , or  $Ce$ ). In our presentation, we have adopted the opposite ordering to reflect the linear order of subject and predicate that is canonically found in English declarative sentences.

The formulation of an explicit model-theory as a system of evaluation of the truth of the formulas of a formal language within a model is due to Alfred Tarski (Tarski 1933, 1944).

Set-theory was developed for the first time as a non-trivial mathematical theory by George Cantor (Cantor 1874) and immediately became a powerful tool in the study of the foundations of mathematics. All mathematical objects, in fact, can be constructed as sets. Hence, set-theory offers a simple, yet comprehensive framework to unify all branches of mathematics. Indeed, Frege's work on the foundations of mathematics relies heavily on set-theory. A significant drawback to set-theory, however, came from Bertrand Russell, who discovered that the theory is ultimately paradoxical (Russell 1903). The reaction to these paradoxes was the development of axiomatic systems, such as the Zermelo-Fraenkel set-theory.

## References

- Cantor, G. (1874). Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen. *Journal für die Reine und Angewandte Mathematik*, 77, 258–262.
- Frege, G. (1879). *Begriffsschrift. Eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle: Verlag von Louis Nebert.
- Russell, B. (1903). *The principles of mathematics*. Cambridge: Cambridge University Press.
- Tarski, A. (1933). The concept of truth in the languages of the deductive sciences. *Prace Towarzystwa Naukowego Warszawskiego, Wydział III Nauk Matematyczno-Fizycznych*, 34(13), 172–198.
- Tarski, A. (1944). The semantic conception of truth: and the foundations of semantics. *Philosophy and Phenomenological Research*, 4(3), 341–376.
- Wittgenstein, L. (1921). Logisch-philosophische abhandlung. In W. Ostwald (Ed.), *Annalen der naturphilosophie* (p. 14). Leipzig: Reinhold.

# Chapter 7

## Meaning and Functions



### Montague's Experiment

Richard Montague was a mathematician and logician who graduated in 1957 from the University of California at Berkeley with a thesis on set-theory and formal languages under the supervision of Alfred Tarski. In the early 1970s, he published three fundamental articles where he experimented with the following question: What if we studied natural languages as formal ones? In the previous chapter, we saw how the meaning of a formal language can be defined as a *homomorphism* between two sets of objects: a set of expressions of the language and a set of objects in a model. What if the meaning of natural languages could be described in the same way? Montague's seminal articles show us that not only it is possible to provide a model-theory of natural language meaning but also that doing so allows us to capture its compositionality on the basis of a general mechanism of composition of meaning.

We begin our illustration of Montague's theory in the most concrete way possible—that is, by providing an actual model-theory of a small group of expressions of English: proper names, intransitive verbs, and the declarative sentences that result from their combination. In the following chapters, we will extend this simple model to some other, slightly more complex expressions.

### Proper Names

Let us begin our illustration of Montague's theory by first focusing on the nominal domain and, in particular, on a specific type of nominal expressions commonly called *proper names*. Proper names are simple expressions that are typically used to refer to individuals. For the purposes of our illustration, let us consider the following proper names:

Kazimir  
 Frida  
 Pablo  
 Alexandra

We can think of proper names as performing the same task performed by lower-case letters in the formal language we devised in the previous chapter. There, lower-case letters were introduced to refer directly to the individual entities in the model. We can think of proper names as performing the same function. Let us therefore design our model of interpretation of the proper names listed above by including in its domain the four individuals Kazimir, Frida, Pablo, and Alexandra, as exemplified in Fig. 7.1.

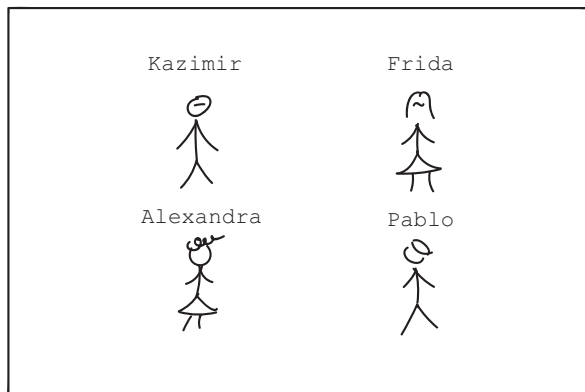
## Intransitive Verbs

We move now to verbs and, in particular, to the class of verbs known as *intransitive verbs*. These are verbs that only require a subject to deliver a declarative sentence. For the purposes of our illustration, let us consider then the following intransitive verbs:

paints  
 sings  
 dances

We can think of intransitive verbs as performing the same function as the upper-case letters in the formal language from the previous chapter. There, we used upper-case letters to refer to sets. We can think of intransitive verbs in the same way. Let us then define the reference of the intransitive verbs “paints”, “sings”, and “dances” as, respectively, the sets of the objects, in the domain of our model, who paint, who sing, and who dance. To make our example more concrete, let us suppose that, in the

**Fig. 7.1** A model of interpretation of the proper names “Kazimir”, “Frida”, “Pablo”, and “Alexandra”





model we are considering, Kazimir, Frida and Alexandra are painting, but Pablo is not. Let us further suppose that Alexandra and Pablo, and only them, are dancing and that Kazimir and Frida, and only them, are singing. In this model, “paints” refers to the set including Kazimir, Frida, and Alexandra, “sings” to the set including Kazimir and Frida, and “dances” to the set including Alexandra and Pablo. This is exemplified in Fig. 7.2.

## Declarative Sentences

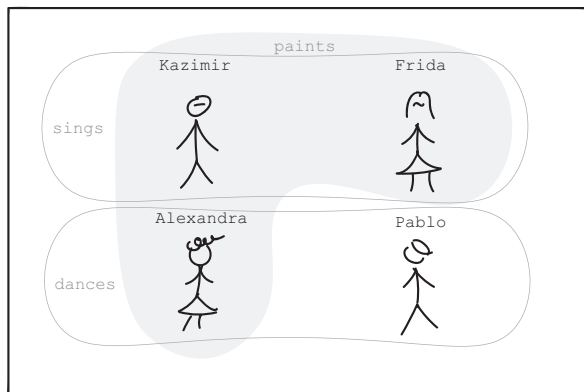
The grammar of English allows us to produce declarative sentences by merging proper names with intransitive verbs. In a phrase-structure grammar, we can represent this grammatical template as in Fig. 7.3, where PN is a label for proper name and IV for intransitive verb. According to this structure, expressions such as “Kazimir dances”, “Frida paints”, and “Alexandra sings” are all grammatical expressions of English.

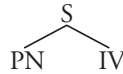
English declarative sentences of the form in Fig. 7.3 are strikingly similar to the combinations of lower-case and upper-case letter that we saw in the formal language in the previous chapter. To interpret them, we introduced a rule that delivers a truth-value, 1 or 0, depending on whether the object referred to by the lower-case letter belongs to the set referred to by the upper-case letter. The expression  $aG$ , as we saw, refers to the truth-value 1 if, and only if,  $a$  belongs to set  $G$ , to 0 otherwise. We can adopt the same strategy for English declarative sentences and define their rule of interpretation as follows:

### Rule of interpretation of English declarative sentences

- a. A sentence  $S$  of the form  $PN + IV$  refers the truth-value 1, if and only if, the object referred to by  $PN$  belongs to the set referred to by  $IV$ ;
- b. A sentence  $S$  of the form  $PN + IV$  refers the truth-value 0, if and only if, the object referred to by  $PN$  does not belong to the set referred to by  $IV$ ;

**Fig. 7.2** A model of interpretation of the intransitive verbs “paints”, “sings”, and “dances”





**Fig. 7.3** In English phrase-structure grammar, merging a proper name (PN) with an intransitive verb (IV) provides a grammatical declarative sentence (S)

According to this rule, a sentence such as “Frida paints” refers to the truth-value 1 if, and only if, the object referred to by “Frida” belongs to the set referred to by “paints” and to the truth-value 0 otherwise.

Notice that the rule is also compositional, in the sense desired by Frege’s principle of compositionality. It derives the meaning of declarative sentences from the meanings of its constituent parts and their mode of grammatical combination. For any proper name and intransitive verb from the lexicon of English, the rule tells you how to interpret the declarative sentence they make together on the sole basis of their individual meanings and the way they are merged together in the sentence. Applied to the sentence “Kazimir dances”, the rule delivers the truth-value 1 if, and only if, Kazimir belongs to the set of those who dance. Applied to the sentence “Alexandra sings”, it delivers the truth-value 1 if, and only if, Alexandra belongs to the set of those who sings. And so on.

This simple model of interpretation is already quite astounding in its potential. For one thing, it provides us with evidence that a formal system of interpretation can be provided for at least a tiny portion of the English language. As limited as this evidence is, it opens the door to an exploration of how much more of English, and, in fact, any natural language, can be explained in this fashion. We should also notice that, despite the limitations in scope that we have just observed, the model we have developed enjoys already a significant degree of generality. As we just observed, our rule of interpretation can be readily generalized to all proper names and intransitive verbs in the lexicon of English, way beyond the few examples we have so far considered. In fact, the rule could be readily generalized to all languages that, like English, can make a declarative sentence by combining a proper name and an intransitive verb.

Montague, however, was not satisfied with what we have achieved so far. By relying on the logical structure of objects and sets, he was able to provide a rule of combination that is so general that it applies to any grammatical operation of Merge.

## Sets, Properties, and Functions

To appreciate Montague’s result, we need first to familiarize ourselves with a new mathematical notion, that of a *function*. Above, we have proposed that intransitive verbs, such as “sings”, “paints”, and “dances”, refer to sets. Sets are collections of objects. In fact, one way to describe them consists in simply enumerating the objects that belong to them, that is, their *members*. In the context of our model, for example,

we can describe the set of those who dance as the set containing Alexandra, Pablo, and no one else. The sets referred to by intransitive verbs, however, are not just arbitrary collections of objects. They are sets of objects that are identified on the basis of a common *property*. Such sets are also called *well-defined* sets. Their members are all the objects that verify a certain condition, and only those. The set referred to by the intransitive verb “dances”, for example, is a well-defined set because it includes all the individuals in the model who satisfy the condition of being dancing, and only those individuals.

Well-defined sets differ from non-well-defined sets in that they can be described not only by listing their members, but also by a function. A function is a mathematical concept we have already encountered in the previous chapter. It is a relation between an input and an output. It takes the input, elaborates it, and delivers the output. The interpretation machine we devised at the end of the previous chapter is a function: It takes a complex expression of the formal language as its input, analyzes whether the expression is true or false in the model, and then delivers a truth-value as its output. The sum operation we all learned at school is also a function: It takes two numbers as its input and delivers a third number, their sum, as its output. Our bodies perform many functions; for example, they take oxygen as their input and deliver energy as their output.

Another thing functions can do is to describe well-defined sets. The set of those who dance can be described by the function that takes an individual  $x$  as its input and delivers a truth-value, 1 or 0, as its output, depending on whether the individual  $x$  dances. In what follows, we will symbolize this function as follows:

(1)  $x \rightarrow 1$  if, and only if,  $x$  dances, 0 if  $x$  does not dance

The symbolization in (1) represents the function as a relation between an input, what appears before the arrow, and an output, what appears after the arrow. The input of the function is represented by an  $x$ —a variable ranging over objects in our model. Simply put, the  $x$  tells us that the input of the function can be anything in our model that is an individual object, such as Kazimir, Frida, Pablo, or Alexandra. The output is a truth-value: 1, in case the value chosen for  $x$  satisfies the property of being dancing; 0, in case it does not. This function describes univocally and unambiguously the set of those objects in the model who dance. If we apply to it the objects of our model, it provides the truth-value 1 for Alexandra and Pablo and the truth-value 0 for Kazimir and Frida. We will call this function the *characteristic function* of the set of those who dance.

As a useful analogy, consider the following example. Imagine having a basket of apples some of which have a diameter of more than 2 inches—let us call them the big apples—and others a diameter of less than 2 inches—let us call them the small apples. Suppose we are asked to single out the small apples. A practical way to do so is to use a hole that is exactly 2 inches wide and try to pass the apples through it. All the apples that pass through the hole have a diameter of less than 2 inches and, therefore, are small apples. All the apples that fail to pass through it have a diameter of more than 2 inches and, therefore, are not small apples. What the hole does is to implement the characteristic function of the set of small apples. Whenever you feed

it with an apple, it tells you whether the apple is or is not a small apple. If you apply it to all the apples in the basket, it delivers the set of small apples.

We can regard the function in (1) as performing a similar task. As the hole singles out the small apples from the set of all the apples in the basket, (1) singles out the individual objects who dance from the set of all objects in the model.

As we can think of the reference of the verb “dances” as either a set or its characteristic function, so we can do with all other intransitive verbs. We can describe the reference of “paints” as the function that maps an object to 1 if it verifies the property of being painting, to 0 if it does not; the reference of “sings” as the function that maps an object to 1 if it verifies the property of being singing, to 0 if it does not.

## Functional Application

The notion of function is central to Montague’s strategy in providing a generalized solution to the challenge of compositionality. Consider again the sentence “Kazimir dances”. The sentence, as we saw, is the product of merging the proper name “Kazimir” with the intransitive verb “dances” to produce the declarative sentence. Above, we have proposed that the proper name “Kazimir” refers to the individual entity Kazimir and that the intransitive verb “dances” refers to the set of dancers. Accordingly, we have stipulated a rule of interpretation that derives the meaning of phrase-structure templates of this sort on the basis of the meaning of its constituent parts. That rule assigns a truth-value to a sentence *S* of the form PN + IV on the basis of whether the reference of PN does or does not belong to the set referred to by IV. According to this rule, the sentence “Kazimir dances” is assigned the truth-value 1 if, and only if, Kazimir belongs to the set of dancers.

In the previous section, we saw that we can also describe the set referred to by “dances” in terms of its characteristic function (1), repeated below:

(1)  $x \rightarrow 1$  if, and only if,  $x$  dances, 0 if  $x$  does not dance

This is a function that takes an individual object  $x$  as its input and delivers a truth-value—1 or 0—as its output, depending on whether  $x$  does or does not satisfy the property of dancing. As we can characterize the meaning of intransitive verbs in terms of a function, so we can reformulate our rule of interpretation of declarative sentences in terms of applications of an input to a function:

### *Functional rule of interpretation of English declarative sentences*

The reference of a declarative sentence *S* of the form PN + IV is the result of applying the object referred to by PN to the function referred to by IV

This rule tells us that the meaning of sentence *S* is the output value we obtain by applying the reference of PN as the input value of the function referred to by IV. Whereas the previous rule assigns a value to *S* on the basis of whether the reference of PN *belongs* to the set referred to by IV, this new rule assigns a value to *S* on the basis of whether the reference of PN *satisfies* the function referred to by IV. This

is why we refer to this rule as a *functional* rule. It derives the meaning of S as the product of the application of an input to a function.

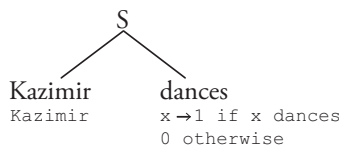
Technically, this new rule and the one we adopted previously are equivalent in that they provide the same truth-values when applied to the same model. The new rule, however, makes it more explicit, from a mathematical perspective, what is the formal process that maps the individual meanings of PN and IV into the truth-value of S. To appreciate this point, let us apply the rule to a concrete example. Consider again the phrase-structure of the declarative sentence “Kazimir dances” along with the meanings of its terminal nodes—the individual Kazimir as the meaning of the proper name “Kazimir” and the characteristic function of the set of dancers in the case of the intransitive verb “dances”. This is exemplified in Fig. 7.4.

The intransitive verb “dances”, on the one hand, refers to a function that takes an individual object as its input and delivers a truth-value as its output. The proper name “Kazimir”, on the other hand, refers to the individual object Kazimir. The rule tells us that the meaning of S is the truth-value obtained by applying the reference of PN, which is the individual Kazimir, as the input value of the function referred to by “dances”, which is the function that delivers the truth-value 1 if, and only if, its input value satisfies the property of dancing. As in the model we are considering Kazimir does not satisfy the property of dancing, the function delivers the truth-value 0. This is the truth-value of S.

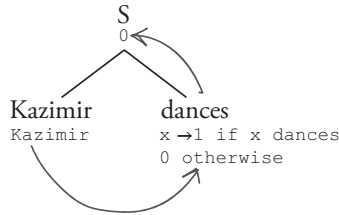
Notice how the rule derives the truth-value of the declarative sentence “Kazimir dances” by applying the individual referred to by “Kazimir” as the input of the function referred to by “dances”, as illustrated in Fig. 7.5.

What is important about the mechanism of application of an input to a function, which in this example we applied to a simple declarative sentence, is that it can be potentially generalized to all other Merge operations in natural language grammar. As we have learned in Chap. 4, natural language grammar is a phrase-structure grammar. Complex expressions are built on the basis of a single operation—Merge—which takes two expressions, complex or simple, and combines them in a larger complex expression—a phrase. All grammatical structures created through Merge share the same general form, illustrated in Fig. 7.6, where  $\alpha$  is the so-called mother node and  $\beta$  and  $\gamma$  are its daughter nodes.

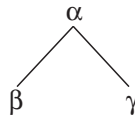
Montague demonstrated that the mechanism of applying an input to a function can be generalized to potentially any grammatical construction of the form in



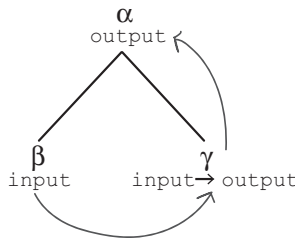
**Fig. 7.4** Phrase-structure of the declarative sentence “Kazimir dances” along with the meanings of its terminal nodes—the individual Kazimir as the meaning of the proper name “Kazimir” and the characteristic function of the set of dancers as the meaning of the intransitive verb “dances”



**Fig. 7.5** The figure illustrates the steps of the procedure that maps the individual meanings of “Kazimir” and “dances” into the truth-value of the sentence. First, the object Kazimir is applied to the function referred to by the verb “dances”. Then, the function assigns a truth-value to S



**Fig. 7.6** A generalized configuration of the structure created by the Merge operation. We will refer to  $\alpha$  as the mother node and to  $\beta$  and  $\gamma$  as the daughter nodes



**Fig. 7.7** A generalized representation of the mechanism of functional application

Fig. 7.6. We will refer to this general mechanism of interpretation of the meaning of complex expressions as *functional application*:

*Functional application*

The reference of a phrase-structure  $\alpha$  of the form  $[\alpha \beta \gamma]$  is the result of applying either  $\beta$  to  $\gamma$  or  $\gamma$  to  $\beta$ .

According to the rule of functional application, whenever we have a phrase  $\alpha$  comprising two daughter nodes  $\beta$  and  $\gamma$ , we can always derive the meaning of  $\alpha$  by either applying the reference of  $\beta$  to the function referred to by  $\gamma$  or, vice versa, by applying the reference of  $\gamma$  to the function referred to by  $\beta$  (Fig. 7.7).

Functional application is, in other words, the semantic counterpart of merge. Whereas merge is the general syntactic principle that creates complex expressions from simple ones, functional application is the general semantic principle that creates complex meanings from simple ones.

At the end of the previous chapter, we saw how we can conceive of the system of interpretation of a formal language as a homomorphism between two sets of objects, a set of expressions of the language and a set of objects in the model of interpretation. With his work, Montague extends this idea to natural language and shows us that we can use it to design a fine-grained homomorphism, which not only mirrors the objects in the two domains but also their *algebras*—that is, the mathematical fabric that holds them together in their respective domains. For each grammatical relation that exists between two expressions in the language, we now have a corresponding relation of functional application between two objects in the model. Grammar and meaning are, in this way, locked one onto the other and the compositionality of natural language meaning is captured by means of a single general combinatorial principle.

This is Montague’s brilliant idea. By expressing the meaning of natural language expressions in terms of reference to objects in a model and exploiting their set-theoretic properties, we can formulate a general mechanism of composition of meaning that mirrors the combinatorial operations of natural language grammar.

## Montague Grammar

In this chapter, we have introduced the core ideas of Montague’s approach to linguistic meaning. His work was seminal to the foundation of the field today known as *model-theoretic natural language semantics*. The name captures the fact that the theory models the meaning of natural language expressions by reference to a model of interpretation. The theory is also known among its practitioners with the friendlier name of *Montague grammar*. The use of the term “grammar” is significant. In effect, the theory provides us with a grammar of meaning—a combinatorial system that constructs complex meanings from simple ones, in harmony with the phrase-structure grammar that characterizes natural language. Central to the functioning of this combinatorial system, as we saw, is the rule of *functional application*. We can regard the mechanism of functional application as the semantic counterpart of the mechanism of merge in syntax. As merge combines simpler grammatical categories into more complex ones, so functional application combines their meanings.

In the decades that followed Montague’s seminal articles, model-theoretic natural language semantics flourished into an independent rich field of research within linguistics. Compositional model-theories have been proposed for a number of grammatical categories and structures of different types and complexity in a variety of different languages. These categories and structures include adjectives, adverbs, relative clauses, complement clauses, classifiers, different classes of nouns, different classes of verbs, negation, pronouns, tense, aspect, measure phrases, comparatives, superlatives, conditionals, interrogatives, exclamatives, evidentials, to mention but a few of the most notable. It is without doubt a great pity that Montague himself—who died violently in March of 1971 in circumstances that remain mysterious—could not witness the fruit of his brilliant work.

## References and Remarks

Montague formulated his ideas on the possibility of a formal analysis of natural language meaning in three articles: “English as a Formal Language” (Montague 1970a); “Universal Grammar” (Montague 1970b); and “The Proper Treatment of Quantification in Ordinary English” (Montague 1973). The last (Montague 1973) is of special importance because it presents an actual formal analysis of a fragment of English, which includes proper names and pronouns, common nouns, quantifiers, tensed intransitive and transitive verbs, sentence taking verbs, prepositions, and verb-modifying and sentence modifying adverbs.

Montague’s 1973 article is known for its difficulty, in large part caused by its “needlessly baroque formalization” (the expression is by Muskens 1995, p. 5). A classic, and more accessible, presentation of its content is provided in David Dowty, Robert E. Wall, and Stanley Peters’s *Introduction to Montague Semantics* (Dowty et al. 1981). Our presentation of Montague’s framework in this and the coming chapters is largely based on his 1973 article, although it simplifies Montague’s original formulation in a number of respects. One important simplification is that, whereas in the system we have presented above natural language expressions are interpreted directly by being assigned objects of reference (or set of objects) in the model, in the framework presented by Montague natural language expressions are first translated in a formal language (that of intensional logic) which is then interpreted in the model (notice, however, that a system of direct interpretation, not mediated by a formal language, was contemplated by Montague in his 1970s article “Universal grammar”; Montague 1970b).

The notion of functional application we have introduced in this chapter is central to Montague’s framework but should be traced back to Frege. In his discussions of the compositionality of language (Frege 1891, 1923), Frege draws a distinction between two types of meanings, saturated and unsaturated. Saturated meanings are meanings that are complete in and by themselves. These are, for example, the meanings of proper names (individual objects) and declarative sentences (truth-values). Unsaturated meanings are meanings that must combine with other meanings in order to produce saturated meanings. These are, for example, the meanings of intransitive verbs; they express properties which, combined with an individual object, produce a truth-value. We also owe to Frege the idea that unsaturated meanings can be expressed as functions and that the process of saturating an unsaturated meaning corresponds to applying a value to a function. What is original about Montague’s approach is that he adopted Frege’s ideas as the basis of a systematic analysis of natural language meaning. Also original about Montague’s contribution is the use of a specific type of logical framework—known as lambda-calculus—to express functions and their combinatoriality systematically. Lambda-calculus is a formal language originally developed by the mathematician Alonzo Church (Church 1932). It owes its name to the fact that it employs the Greek letter  $\lambda$  (lambda) to express functions.

The generalized mechanism of functional application we have presented in this chapter—a single rule of composition that applies to all phrase-structures—does not



appear in such a general form in Montague's original writings. It is true that Montague adopts the mechanism of functional application as the mode of meaning composition that is common to all the grammatical constructions that appear in his fragment of English. Nonetheless, when formulating his framework, he provides specific rules of composition for the different grammatical constructions. This approach was later revised by Klein and Sag (1985), who formulated the mechanism of functional application in the generalized form we have presented in this chapter (an approach known as *type-theory*).

Montague's writings were without doubt amongst the most influential in fostering the formal approach to natural language meaning. Yet, they were not the only ones. Similar ideas were promoted by other scholars at around the same time. Relevant examples are David Lewis's article "General Semantics" (Lewis 1970), Max Cresswell's book *Logics and Languages* (Cresswell 1973) and Terence Parsons's thesis manuscript "An Outline of a Semantics of English" (Parsons 1972).

A pivotal figure in introducing Montague's ideas to the forum of theoretical linguistics was that of Barbara H. Partee. As perhaps the only linguist at Montague's time who was knowledgeable in both formal syntax and formal semantics, Partee was the first to recognize the possibility of reconciling the two aspects of natural language—form and meaning—within a unifying framework. Partee contributed to the development of contemporary formal semantics through her many publications (among the most notable Partee 1973, 1974, 1975, 1976, 1980, 1981, 1986, 1996), as well as by mentoring the graduate work of more than one generation of students and scholars.

Today, several textbooks are available to those who wish to familiarize themselves with the formal approach to natural language semantics. Dowty, Wall, and Peters's *Introduction to Montague Semantics* (Dowty et al. 1981) is a classic introduction to Montague's theory, its formal foundations, and its applications. *Meaning and grammar: An introduction to semantics* by Gennaro Chierchia and Sally McConnell-Ginet (Chierchia and McConnell-Ginet 2000) and *Semantics in Generative Grammar* by Irene Heim and Angelika Kratzer (Heim and Kratzer 1998) are amongst the most widely adopted introductions to formal semantics across linguistics departments around the world. An extremely accessible and enjoyable introduction to formal semantics is Emmon Bach's *Informal Lectures in Formal Semantics* (Bach 1989). A number of specialized scientific journals have established themselves in the field of formal semantics and publish with regularity the work of scholars in this field. Some of the most known are *Journal of Semantics*, *Natural Language Semantics*, and *Semantics and Pragmatics*.

Finally, a note of caution is in order when talking about transitivity. A correct analysis of the notion of transitivity requires drawing a distinction between syntactic and semantic transitivity. Syntactic transitivity concerns the number of arguments a verb requires in order to provide a grammatical sentence. An intransitive verb (such as "dance") requires only one argument (as in the sentence "Kazimir dances"). A transitive verb (such as "admire") requires two arguments (as in "Kazimir admires Frida"). A ditransitive verb (such as "introduce") requires three (as in "Kazimir introduced Frida to Pablo"). Semantic transitivity concerns the

number of participants that are involved in the eventuality described by the verb. It is easy to overlook this distinction and conflate the two notions. However, there are a number of observable cases in which the two notions do not align. Consider, as an example, the verb “paint”. As we have observed, the verb can be used as an intransitive verb. This is demonstrated by the fact that the sentence “Kazimir paints” is perfectly grammatical. Semantically, however, any act of painting is bound to include an object that is painted. In fact, the same verb can also be used as a transitive verb (as in the sentences “Kazimir paints a canvas” or “Kazimir paints a square”). For the sake of simplicity, we will maintain throughout the book that the syntactic transitivity of a verb determines exactly the number of semantic arguments to be considered when computing its interpretation. Although this simplification will not affect in any significant way our arguments and conclusions, it is nonetheless an important concern for anyone interested in the mapping from syntactic structure into meaning.

## References

- Bach, E. (1989). *Informal lectures on formal semantics*. Albany: SUNY Press.
- Chierchia, G., & McConnell-Ginet, S. (2000). *Meaning and grammar: An introduction to semantics*. Cambridge: MIT press.
- Church, A. (1932). A set of postulates for the foundation of logic. *Annals of mathematics*, 33, 346–366.
- Cresswell, M. J. (1973). *Logics and languages*. London: Methuen.
- Dowty, D., Wall, R. E., & Peters, S. (1981). *Introduction to montague semantics*. Dordrecht: D. Reidel.
- Frege, G. (1891). *Funktion und Begriff. Vortrag, gehalten in der Sitzung vom 9. Januar 1891 der Jenaischen Gesellschaft für Medizin und Naturwissenschaft*. Jena: Hermann Pohle.
- Frege, G. (1923). Logische Untersuchungen. Dritter Teil: Gedankengefüge. *Beiträge zur Philosophie des deutschen Idealismus*, 3, 36–51.
- Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Oxford: Blackwell.
- Klein, E., & Sag, I. A. (1985). Type-driven translation. *Linguistics and Philosophy*, 8(2), 63–201.
- Lewis, D. (1970). General semantics. *Synthese*, 22(1-2), 18–67.
- Montague, R. (1970a). English as a formal language. In B. Visentini et al. (Eds.), *Linguaggi nella società e nella tecnica* (pp. 189–224). Milan: Edizioni di Comunità.
- Montague, R. (1970b). Universal grammar. *Theoria*, 36(3), 373–398.
- Montague, R. (1973). The proper treatment of quantification in ordinary English. In *Approaches to natural language* (pp. 221–242). Dordrecht: D. Reidel.
- Muskens, R. (1995). *Meaning and partiality*. Stanford, CA: CSLI/FoLLI.
- Parsons, T. (1972). *An outline of a semantics of English*. Amherst: Ms. University of Massachusetts.
- Partee, B. (1973). Some transformational extensions of Montague grammar. *Journal of Philosophical Logic*, 2(4), 509–534.
- Partee, B. (1974). Opacity and scope. In M. Munitz & P. Unger (Eds.), *Semantics and philosophy*. New York: NYU Press.
- Partee, B. (1975). Deletion and variable binding. In *Formal semantics of natural languages* (pp. 16–34). Cambridge: Cambridge University Press.
- Partee, B. (Ed.). (1976). *Montague grammar*. New York: Academic Press.
- Partee, B. (1980). Semantics – mathematics or psychology? In E. Bäuerle & A. von Stechow (Eds.), *Semantics from different points of view* (pp. 1–14). Berlin: Springer-Verlag.

- Partee, B. (1981). Montague grammar, mental representations, and reality. In S. Kanger & S. Öhman (Eds.), *Philosophy and grammar* (pp. 59–78). Dordrecht: D. Reidel.
- Partee, B. (1986). Noun phrase interpretation and type-shifting principles. In J. Groenendijk, D. de Jongh, & M. Stockhof (Eds.), *Studies in discourse representation theory and the theory of generalized quantifiers* (pp. 115–143). Dordrecht: Foris.
- Partee, B. (1996). The development of formal semantics in linguistic theory. In S. Lappin (Ed.), *The handbook of contemporary semantic theory* (pp. 11–38). Oxford: Blackwell.

# Chapter 8

## Meaning and Relations



### Transitive Verbs

In the previous chapter, we have familiarized ourselves with the core workings of Montague's formal approach to natural language meaning. We have done so by providing a model-theory of the meaning of proper names, intransitive verbs, and the declarative sentences that result from their combination. We will dedicate the remainder of part I to extending this model to a few more grammatical categories and structures of higher complexity. The model of natural language meaning that will result from this exercise will provide us with a framework for formulating and assessing a number of more fundamental questions about the nature of meaning—a task that we will undertake in parts II and III.

In this chapter, we will extend our current model to transitive verbs. Transitive verbs are verbs such as “love”, “hate”, “admire”, “despise”, “help”, and “reject”. Differently from intransitive verbs, these verbs must combine with two names—or noun phrases—in order to produce a grammatical sentence. So, for example, the complex expression “Pablo admires Alexandra” is a grammatical sentence of English whereas “Pablo admires” is not. In the previous chapter, we have seen that intransitive verbs express properties of objects. In this chapter, we will see that transitive verbs can be regarded as expressing *relations* between objects. We will henceforth begin by familiarizing ourselves with the notion of relation.

### Relations in Model-Theory

To best appreciate the notion of relation from the perspective of a model-theory, let us consider again the model we introduced in Chap. 6. As we saw, the model features a number of geometric figures organized on the basis of two properties, shape

and color. When studying the objects in a model, however, we may also want to consider the relations that hold between them. By looking more carefully at the figures in our model, we find, for examples, that some figures can be inscribed in others. Circle  $b$  can be inscribed in square  $d$ , square  $f$  can be inscribed in triangle  $i$ , and triangle  $j$  can be inscribed in circle  $a$ , as shown in Fig. 8.1.

In a model, properties correspond to sets of objects—the set of all objects that satisfy the property. The property of being grey, for example, corresponds to the set of all objects in the model that are grey. It is possible to express relations in a similar manner. Suppose we want to make an inventory of which figures can be inscribed in which figures in the model. A simple way to do so is by making a list of *pairs*: all the pairs of figures such that the first figure in each pair can be inscribed in the second.

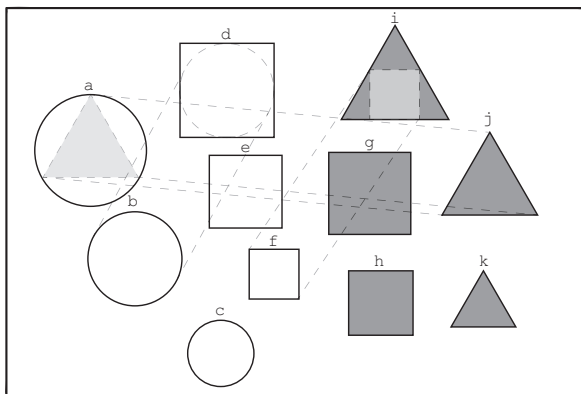
- (1)  $\langle b, d \rangle$   
 $\langle f, i \rangle$   
 $\langle j, a \rangle$

Notice that the pairs listed in (1) are all enclosed between angled brackets. This is a typographical convention to indicate that the pairs are *ordered* pairs, that is, pairs where it matters which is the first and which is the second element in the pair. Each pair in the list tells us that the first figure in the pair can be inscribed in the second. For example, the pair  $\langle b, d \rangle$  tells us that figure  $b$  can be inscribed in figure  $d$ —not that figure  $d$  can be inscribed in figure  $b$ .

The set of pairs in (1) offers a complete description of the relation of inscription in our model. If we want to check whether a figure can be inscribed in another figure in our model, all we need to do is check whether the pair composed by the two figures belongs to the list in (1). If we want to know, for example, whether figure  $c$  can be inscribed in figure  $k$ , all we need to do is look for the pair  $\langle c, k \rangle$  in the list in (1). Since the pair does not appear in the list, we conclude that  $c$  cannot be inscribed in figure  $k$ .

This quick exercise suggests that, if a property can be regarded as a set of objects—all objects that satisfy the property—a relation can be regarded as a set of ordered pairs—all pairs of objects that satisfy the relation.

**Fig. 8.1** A representation of the model of geometric figures from the previous chapters that also shows which figures can be inscribed in which figures: Circle  $b$  can be inscribed in square  $d$ , square  $f$  can be inscribed in triangle  $i$ , and triangle  $j$  can be inscribed in circle  $a$



## Relations in Formal Language

Having an understanding of relations in terms of sets, we can now easily express them in a formal language. We can provide a concrete example by extending the formal language we developed in Chap. 6. To do so, we must first introduce a new category of symbols in the lexicon of our formal language. The formal language we developed in Chap. 6 included two types of expressions: lower-case letters, referring to objects, and upper-case letters, referring to sets. We will now use *bold upper-case letters* to refer to sets of ordered pairs, that is, relations. In particular, we will use the letter **I** to refer to the set of pairs in (1).

### Lexicon

- a. Lower-case letters: *a, b, c, d, e, f, g, h, i, j*, and *k*;
- b. Upper-case letters: *C, S, T, B*, and *W*.
- c. Bold upper-case letters: **I**

We must then extend the grammar of the language by adding a new grammatical rule governing the use of bold upper-case letters:

### Grammar

- a. Any lower-case letter followed by any upper-case letter is a grammatical combination;
- b. Any lower-case letter followed by bold upper-case letter followed by a lower-case letter is a grammatical combination;
- c. Any other combination is ungrammatical.

The relevant addition to the grammatical rules we had already set in Chap. 6 is (b). This rule allows us to combine a bold upper-case letter with two lower-case letters, one immediately preceding it and one immediately following it. Hence, *aIb*, *kId*, and *eIh* are all grammatical expressions of the language, whereas *Iad*, *Ifhj*, and *SIBa* are not.

Finally, we must formulate a rule of interpretation for the complex expressions produced by this new grammatical rule:

### Interpretation

- a. A complex expression of the form *lower-case letter + upper-case letter* is assigned the truth-value 1 if, and only if, the object referred to by the lower-case letter belongs to the set referred to by the upper-case letter.
- b. A complex expression of the form *lower-case letter + upper-case letter* is assigned the truth-value 0 if, and only if, the object referred to by the lower-case letter does not belong to the set referred to by the upper-case letter.
- c. A complex expression of the form *lower-case letter + bold upper-case letter + lower-case letter* is assigned the truth-value 1 if, and only if, the pair including the referent of the first lower-case letter and of the second lower-case letter belongs to the set of pairs referred to by the bold upper-case letter.
- d. A complex expression of the form *lower-case letter + bold upper-case letter + lower-case letter* is assigned the truth-value 0 if, and only if, the pair including

the referent of the first lower-case letter and of the second lower-case letter does not belong to the set of pairs referred to by the bold upper-case letter.

The new rules are (c) and (d). They tell us that an expression such as  $bId$  refers to the truth-value 1 if the pair of figures  $\langle b, d \rangle$  is a member of the set denoted by  $I$ , to the truth-value 0 if it does not. Above, we introduced the symbol  $I$  to refer to the set of pairs (1), repeated below for the reader's convenience.

- (1)  $\langle b, d \rangle$   
 $\langle f, i \rangle$   
 $\langle j, a \rangle$

This is the set of all pairs of figures such that the first figure can be inscribed in the second. Notice that the pair  $\langle b, d \rangle$  is, in fact, a member of the set. Hence, given our rules of interpretation, the expression  $bId$  is assigned the truth-value 1. Informally, our new rules tell us that the expression  $bId$  is true if, and only if, figure  $b$  can be inscribed in figure  $d$ .

## Relations in Natural Language

In English, as in many other natural languages, relations between objects are typically expressed by means of *transitive verbs*. Sentence (2), for example, is intuitively understood as expressing a relation of admiration between Pablo and Alexandra. Similarly, sentence (3) is understood as expressing a relation of despise between Frida and Kazimir.

- (2) Pablo admires Alexandra  
 (3) Frida despises Kazimir

These sentences are similar to the expressions we have provided in our formal language to express relations and are, therefore, passible of a similar analysis. For concreteness' sake, let us assume that, in the model we are considering, Frida admires Pablo and Pablo admires Alexandra but Kazimir despises Frida and Alexandra despises Pablo. Let us then take the set of pairs in (4) to be the set referred to by the verb "admires" and the set of pairs in (5) to be the set referred to by "despises".

- (4)  $\langle \text{Frida, Pablo} \rangle$   
 $\langle \text{Pablo, Alexandra} \rangle$   
 (5)  $\langle \text{Kazimir, Frida} \rangle$   
 $\langle \text{Alexandra, Pablo} \rangle$

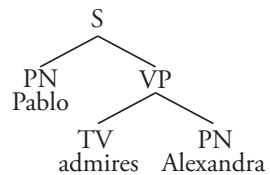
The set in (4) is the set of all pairs of objects in our model such as the first object in the pair admires the second; (5) is the set of all pairs of objects in our model such as the first object despises the second.

On a parallel with the formal language, we can then establish a rule of interpretation according to which a sentence of the form *proper name + transitive verb + proper name* is assigned the truth-value 1 if, and only if, the pair of objects referred to by the first and the second proper names in the sentence belongs to the set referred to by the transitive verb. According to such rule, the sentence “Pablo admires Alexandra” is true in the model if, and only if, the pair  $\langle \textit{Pablo}, \textit{Alexandra} \rangle$  belongs to the set referred to by the verb “admires”. Given that, as we can see in (1), the pair  $\langle \textit{Pablo}, \textit{Alexandra} \rangle$  does belong to the set of pairs referred to by the verb “admires”, the sentence is assigned the truth-value 1.

This rule does the job and demonstrates that transitive verbs are passible of a formal analysis. The rule has, nonetheless, a problem. It is not compositional. It fails to capture how the meaning of the sentence is constructed on the basis of its grammatical structure, as demanded by the principle of compositionality. Syntacticians tell us that the sentence “Pablo admires Alexandra” is produced by two successive applications of the Merge operation, as exemplified in Fig. 8.2. First, the transitive verb “admires” is merged with the proper name “Alexandra” to form the verb phrase (VP) “admires Alexandra” and then the verb phrase “admires Alexandra” is merged with the proper name “Pablo” to form the sentence (S).

The principle of compositionality demands that the meaning of the sentence be derived in the same fashion. First, the meanings of “admires” and “Alexandra” must combine to provide the meaning of the verb phrase “admires Alexandra”. Then, the meaning of the verb-phrase “admires Alexandra” must combine with the meaning of “Pablo” to deliver the meaning of the whole declarative sentence. To achieve this result within Montague’s framework, we must find a way to translate the sets of ordered pairs referred to by transitive verbs into mathematical functions. Doing so would allow us to derive the meaning of the declarative sentences constructed from transitive verbs on the basis of the same general principle of functional application we introduced in the previous chapter. This, however, is no easy task. To begin with, we must find a way to translate sets of ordered pairs, such as those discussed above, into functions. Furthermore, the type of functions we need for transitive verbs cannot be simple mappings from an input into an output, as in the case of intransitive verbs. What we really need are functions that take two different inputs in two different steps in order to deliver a final output value. For example, in the case of the sentence “Pablo admires Alexandra”, what is needed is a function that takes the individual Alexandra as its first input value, then takes the individual Pablo as its second input value, and only then delivers the truth value 1 as its value if, and only if, Pablo admires Alexandra. Fortunately, there is a way to provide such functions, which is based on the work of the mathematicians Haskell B. Curry and Moses

**Fig. 8.2** The phrase-structure of the sentence “Pablo admires Alexandra”





Ilyich Schönfinkel. The next section will be devoted to illustrating this type of functions. We will then see how they can be applied to transitive verbs to provide a fully compositional account of their meaning.

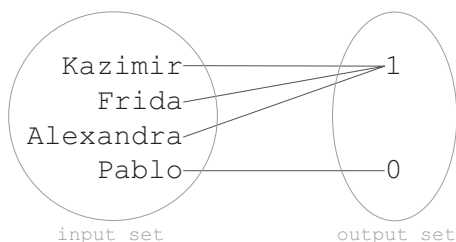
## Functions of Functions

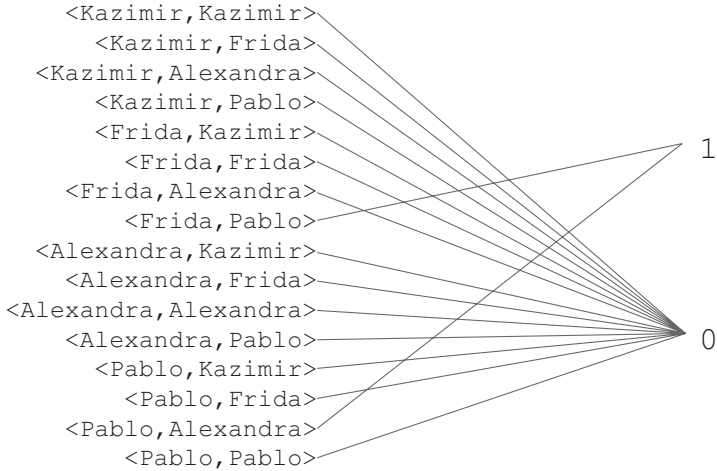
In the previous chapter, we saw that well-defined sets can be described by a characteristic function. The set of painters, for example, can be described by a corresponding function taking an object as its input and delivering the truth-value 1 if the object paints and the truth-value 0 if it does not. In effect, what the function does in the context of a model of objects is to associate each object in the model with a truth-value—1 or 0—depending on whether the object does or does not meet the relevant property. For example, the content of the function denoted by the verb “paints” in the context of our model of interpretation can be represented as in Fig. 8.3. The figure presents the function as a relation between two sets: an input set—the set of all objects in the model that are possible inputs of the function—and an output set—the set of all objects that are possible outputs of the function. The lines between the two sets tell us which input delivers which output. The figure tells us, for example, that Frida is a painter—because, when selected as the input of the function, the function delivers as its output the truth-value 1. The figure also tells us that Pablo is not a painter—because, when chosen as the input of the same function, the function delivers the truth-value 0.

Importantly, the same type of representation can be produced for relations. As relations can also be analyzed as sets, so they can be described by a characteristic function. The only peculiar aspect of relations is that their input set consists of pairs of objects, rather than individual objects. We can therefore represent a relation, such as the one denoted by the transitive verb “admires”, as in Fig. 8.4.

The representation in Fig. 8.4 connects all the pairs of objects such that the first object admires the second to the output value 1 and all the pairs such that the first object does not admire the second to the output value 0. We gather from the figure that, for example, Pablo admires Alexandra. This is because the pair  $\langle \text{Pablo}, \text{Alexandra} \rangle$  is mapped into the value 1. We also understand that Alexandra does not

**Fig. 8.3** The function denoted by “paints” represented, in the context of our model of interpretation, as a relation between an input set and an output set



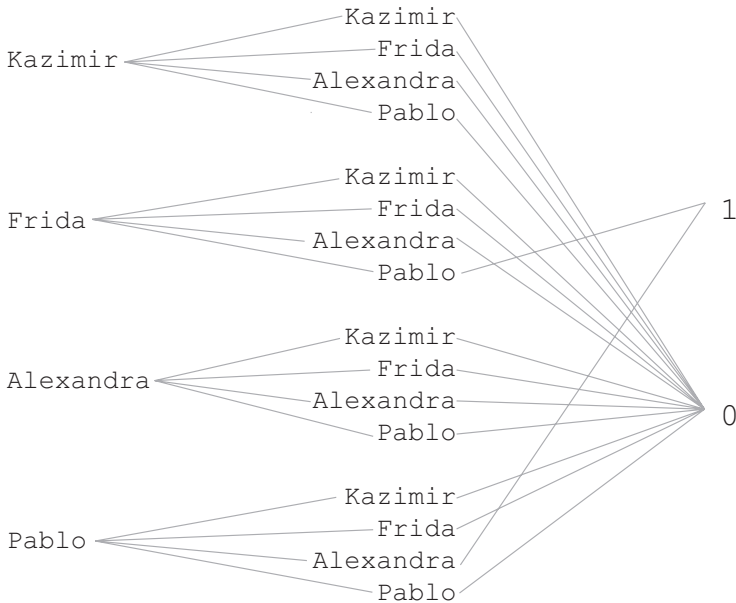


**Fig. 8.4** The function denoted by “admires” represented, in the context of our model of interpretation, as a relation between an input set and an output set. Notice that, in this case, the input set is a set of ordered pairs

admire Pablo. This is because the pair *<Alexandra, Pablo>* is mapped into the value 0.

The function depicted in Fig. 8.4 has a mathematical property that is especially useful in the context of our enterprise. It is possible of a type of transformation known as *currying*—after the mathematician Haskell Curry—that translates a function from ordered pair of objects into a sequence of two functions from individual objects. The easiest way to understand how currying works is to see it in action. There are two ways in which we can apply the currying transformation to the function in Fig. 8.4. Both consist in splitting up the input set—which in Fig. 8.4 is a set of pairs of objects—into two distinct sets of individual objects. The first type of application is represented in Fig. 8.5.

Figure 8.5 describes the relation referred to by the verb “admires” as a relation between three, rather than two sets. The lines represent the mapping from a first individual entity into a second individual entity into, finally, a truth-value. The first input set is the set of potential *admirers*. The second input set is the set of potentially *admired* individuals. The output set is, as before, the set of truth-values. The function in Fig. 8.5 is entirely equivalent to the one in Fig. 8.4. It maps us to the truth-value 1 if and only if the individual we pick in the first set admires the individual we pick in the second set. Suppose that we want to find whether Pablo admires Alexandra. What we need to do is point our finger to the individual Pablo in the first set, follow the line that connects it to Alexandra, in the second set, and then follow the line that connects Alexandra to a truth-value. As the line takes us to the truth-value 1, we conclude that Pablo admires Alexandra. Suppose, instead, that we want to find whether Alexandra admires Pablo. What we need to do is point our finger to the individual Alexandra in the first set, follow the line that connects it to Pablo in



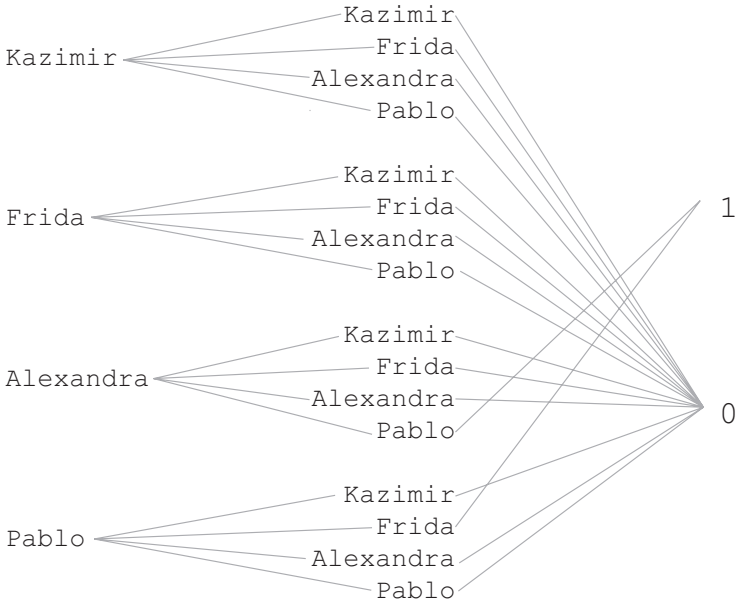
**Fig. 8.5** First application of the currying transformation to the function in Fig. 8.4

the second set, and finally follow the line that connects Pablo to a truth-value. As the line takes us to the truth-value 0, we conclude that Alexandra does not admire Pablo. You can easily experiment with other pairs of individuals and verify that the functions in Figs. 8.4 and 8.5 provide the exact same results and are, therefore, equivalent.

As anticipated, there is a second way in which we can apply the currying transformation to the function in Fig. 8.4. This is provided in Fig. 8.6. In this case the first set is the set of potentially admired individuals, whereas the second input set is the set of potential admirers. Hence, to verify whether Pablo admires Alexandra, we must first point to Alexandra—the potentially admired individual—then follow the line that connects Alexandra to Pablo—the potential admirer—and finally follow the line that takes us to a truth-value. Since we end up pointing at the truth-value 1, we conclude that Pablo admires Alexandra. Similarly, to verify whether Alexandra admires Pablo, we must first point to Pablo—the potentially admired individual—then follow the line that connects Pablo to Alexandra—the potential admirer—and finally follow the line that takes us to a truth-value. As we end up pointing at the truth-value 0, we conclude that Alexandra does not admire Pablo. In this case as well, you can easily experiment with other pairs of individuals to verify that the functions in Figs. 8.4, 8.5, and 8.6 provide all the very same result.

The lesson we learn is that we can always transform a function of the form symbolized in (6a) into equivalent functions of the form in (6b) and (6c).

- (6a)  $\langle x, y \rangle \rightarrow 1$  if  $x$  admires  $y$ , 0 otherwise
- (6b)  $x \rightarrow (y \rightarrow 1$  if  $x$  admires  $y$ , 0 otherwise)
- (6c)  $y \rightarrow (x \rightarrow 1$  if  $x$  admires  $y$ , 0 otherwise)



**Fig. 8.6** Second application of the currying transformation to the function in Fig. 8.4

The symbolization in (6a) describes a function that maps an ordered pair of individual objects  $x$  and  $y$  into the value 1 if, and only if,  $x$  admires  $y$ ; to 0 otherwise. The symbolization in (6b) describes a function that maps a first individual entity  $x$  into another function that maps a second individual entity  $y$  into the value 1 if, and only if, the  $x$  admires  $y$ , to 0 otherwise. The symbolization in (6c) is the function that maps a first individual entity  $y$  into another function that maps a second individual entity  $x$  into the value 1 if, and only if,  $x$  admires  $y$ , to 0 otherwise.

Whereas the three functions are perfectly identical in the result they deliver, they differ in how they produce such result. In the case of (6a), we must feed  $x$  and  $y$  to the function as one single input—the ordered pair  $\langle x, y \rangle$ . In the case of (6b) and (6c), conversely, we feed the values  $x$  and  $y$  one at the time, in two successive steps. In the case of (6b), we first apply  $x$ , then  $y$ , and then we obtain the resulting truth-value. In the case of (6c), we first apply  $y$ , then  $x$ , and then we obtain the truth-value.

We will see in the following section that the function in (6c) is precisely what we need in order to provide a fully compositional account of the meaning of transitive verbs.

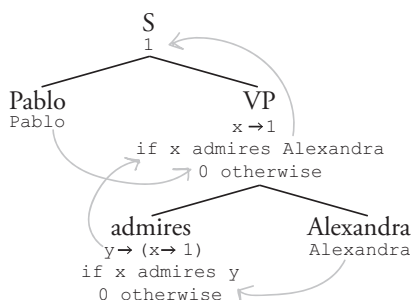
## Relations in Montague's Semantics

Consider again the sentence “Pablo admires Alexandra” together with its phrase-structure, as represented in Fig. 8.2. Assume the following meanings for the simple expressions occurring in it. In the case of “Pablo” and “Alexandra”, assume, as we have done so far, that they refer to the individuals Pablo and Alexandra, respectively. In the case of the verb “admires”, let us now assume that it refers to the characteristic function (6c), repeated below.

$$(6c) \quad y \rightarrow (x \rightarrow 1 \text{ if } x \text{ admires } y, 0 \text{ otherwise})$$

As we saw in the previous section, (6c) is the function that takes an object  $y$  as its input and delivers a second function as its output, which takes an individual  $x$  as its input and delivers as its output the value 1 only in the case in which  $x$  admires  $y$  in the model of evaluation of the function.

Adopting these meanings allows us to derive the truth-value of the whole sentence in a fully compositional fashion, by combining the meanings of its constituent in the same order in which they are merged in the phrase-structure of the sentence. This process is exemplified in Fig. 8.7. As detailed in the figure, the meaning of the sentence is produced in a step-by-step fashion. The first step consists in applying the reference of “Alexandra” to the function corresponding to the verb “admires”. This, as we saw, is a function that takes an individual as its input—in this case Alexandra—and delivers another function as its output. When we apply Alexandra to this function, we obtain, as a result, a function that maps an object  $x$  into the truth-value 1 if  $x$  admires Alexandra, to 0 otherwise. This is the meaning of the VP “admires Alexandra”. The second step consists in applying the reference of “Pablo” to this second function. Since, in the model we are considering, Pablo does admire Alexandra, the function delivers the truth-value 1 as the meaning of whole sentence.



**Fig. 8.7** A fully compositional derivation of the meaning of the sentence “Pablo admires Alexandra”. In the first step, Alexandra is applied to the function  $y \rightarrow (x \rightarrow 1$  if  $x$  admires  $y$ , 0 otherwise), which is the reference of the verb “admires”. This application delivers the function  $x \rightarrow 1$  if  $x$  admires Alexandra, 0 otherwise; this function is the reference of the VP “admires Alexandra”. In the second step, Pablo is applied to the function  $x \rightarrow 1$  if  $x$  admires Alexandra, 0 otherwise. This application delivers, in the model of interpretation we are adopting, the truth-value 1, which is the reference of the sentence

By relying on the currying transformation, we have met our challenge. We have provided a fully compositional account of English declarative sentences with transitive verbs, where the meanings of the different constituents of the sentence combine with one another in the order dictated by their grammatical structure. The unifying principle is, once again, that of functional application, whereby the meaning of a complex grammatical expression is the result of the application of an input to a function. Fundamental to achieving this result was understanding relations in terms of functions of functions. This exercise gave us a first glimpse into how the formal properties of Montague's framework can be exploited to solve problems of compositionality of higher and higher complexity. In the next chapter, we will apply a similar strategy to provide a compositional account of yet another class of expressions, complex noun phrases resulting from the combination of a determiner and a noun.

## References and Remarks

The process of currying owes its name to the mathematician Haskell B. Curry although, according to *The Development of Logic* by William Calvert Kneale and Martha Kneale (Kneale and Kneale 1962), the mathematician Moses Ilyich Schönfinkel should be credited for it (Schönfinkel 1924). Irene Heim and Angelika Kratzer adopt this practice in their textbook (Heim and Kratzer 1998) and, accordingly, refer to the process as Schönfinkelization.

## References

- Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Oxford: Blackwell.  
Kneale, W. C., & Kneale, M. (1962). *The development of logic*. Oxford: Clarendon Press.  
Schönfinkel, M. (1924). Über die Bausteine der mathematischen Logik. *Mathematische Annalen*, 92(3-4), 305–316.

## Chapter 9

# Meaning and Quantification



### Determiners

The complex functions we have employed in the previous paragraph to explain transitive verbs apply, in fact, to a number of other grammatical structures. In this section, we will see how they can be extended to complex noun phrases. The only exemplars of nominal expressions we have considered so far are proper names, such as “Kazimir” and “Alexandra”. Proper names are the simplest type of nominal expressions. They are able to combine with verbs—transitive and intransitive alike—to form sentences. Proper names, however, are not the only expressions capable of performing this function. Consider the sentences in (1). The subjects of these sentences are not simple proper names but combinations of a *determiner*—“every”, “no”, or “a”—and a *common noun*—“dancer”, “singer”, or “painter”.

- (1a) Every dancer paints
- (1b) No singer dances
- (1c) Some painter dances

To be sure, the class of expressions syntacticians refer to as determiners is a very rich one and, in fact, different classes of determiners have often received different semantic treatments. In this chapter, we will discuss one of the most general approaches to the meaning of determiners, which attempts at providing a common framework for the interpretation of all determiners. The only exception we will make to this somewhat simplified picture will concern the definite article “the”, for which we will provide a separate discussion in the next chapter. The challenge we will address in this chapter is how to derive the meaning of sentences such as those in (1) in a compositional fashion—that is, on the basis of the individual meanings of their constituent parts and the grammatical structure that holds them together. We

will address this challenge by first considering the meaning conveyed by these sentences and how it can be represented in a model. We will then see how this meaning can be constructed compositionally.

## Quantification in Model-Theory

Intuitively, the sentences in (1) can all be understood as expressing relations, although not relations between individual objects, as those conveyed by the transitive verbs we discussed in the previous chapter, but logical relations between properties. Sentence (1a) tells us that the property of being a dancer *entails* the property of being a painter: If one is a dancer, then one is also a painter. Sentence (1b) tells us that the property of singing and that of dancing stand in a relation of *contradiction*: There is no one that both sings and dances. Sentence (1c) expresses the *conjunction* of the properties of painting and dancing: There is at least one individual in the model that both paints and dances.

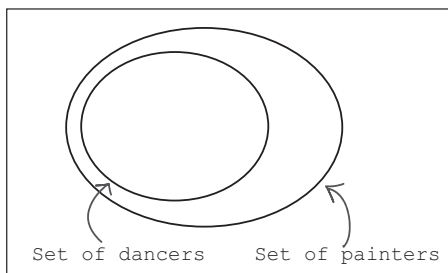
Importantly from our perspective, each of these logical relations can be described in a model of objects by a corresponding relation between sets. Sentence (1a) is true in a model if, and only if, the set of dancers is a *subset* of the set of painters, as depicted in Fig. 9.1. Whenever this configuration between the two sets is met, any individual object that belongs to the set of dancers, also automatically belongs to the set of painters. This is the configuration that makes sentence (1a) true.

Sentence (1b) is true if, and only if, the *intersection* between the set of singers and the set of dancers is *empty*, as depicted in Fig. 9.2. All models that are characterized by such a relation between the two sets are models where (1b) is true.

Finally, sentence (1c) is true if, and only if the *intersection* between the set of painters and the set of dancers is *non-empty*, as depicted in Fig. 9.3. A model where there is at least one object that belongs to the intersection between the two sets is a model that makes sentence (1c) true.

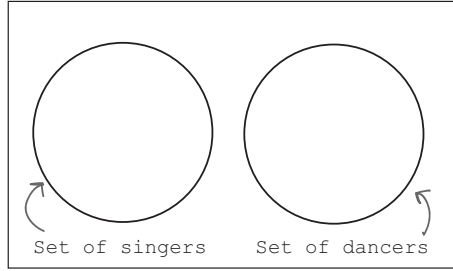
Consider, as a useful exercise, the truth-values the three sentences receive in the context of the model of interpretation we have used in the previous chapters, here reproduced in Fig. 9.4. In this model, we find that sentence (1a) is false: It is not true that the set of those who dance is a subset of the set of those who paint. Pablo, who belongs to the set of dancers, is not included in the set of painters. Hence, it is not

**Fig. 9.1** The set-theoretic relation between the set of dancers and the set of painters expressed by the sentence “every dancer paints”

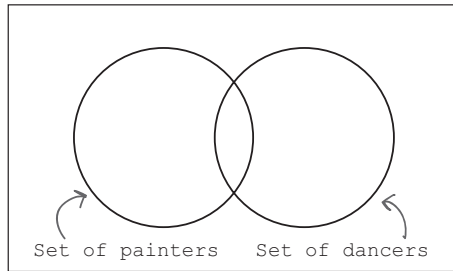




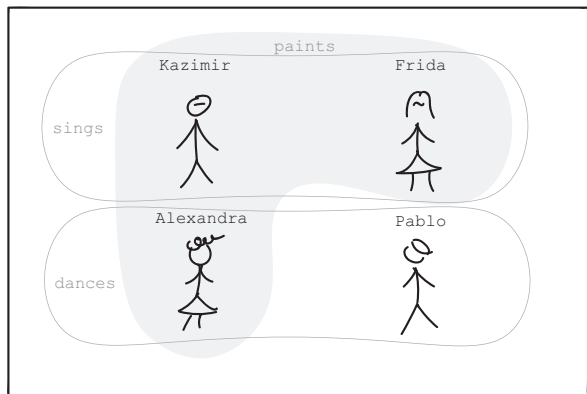
**Fig. 9.2** The set-theoretic relation between the set of dancers and the set of singers expressed by the sentence “No singer dances”



**Fig. 9.3** The set-theoretic relation between the set of dancers and the set of painters expressed by the sentence “Some painter dances”



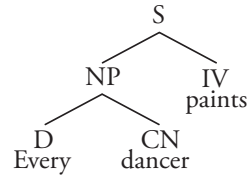
**Fig. 9.4** The model of interpretation from Chap. 7, representing the four individuals Kazimir, Frida, Pablo, and Alexandra and the three properties of painting, singing, and dancing



true that every dancer paints. Conversely, we find that sentence (1b) is true. As there is no intersection between singers and dancers, it is true that no singer dances. In this model, sentence (1c) is also true. As the intersection between painters and dancers is non-empty—it includes Alexandra—it is true that some painter dances.

The lesson we draw from this exercise is that the sentences in (1) express relations between properties which, in turn, can be represented in a model of objects as relations between sets. From now on we will refer to relations of this sort as relations of *quantification*. In the next section, we will see how relations of quantification can be expressed compositionally.

**Fig. 9.5** Phrase-structure of the sentence “every dancer paints”



## Quantification in Montague Grammar

As we did in the previous chapter, we can derive the meaning of the sentences in (1) compositionally by relying on the principle of functional application and the currying transformation. Let us begin by considering the grammatical structure of these sentences. For the time being, we will use (1a) as our chief example and will then extend the analysis to the other sentences. As exemplified in Fig. 9.5, sentence (1a) is constructed by first merging the determiner (D) “every” with the common noun (CN) “dancer” to produce the noun phrase (NP) “every dancer” and by then merging this noun phrase with the intransitive verb “paints” to produce the sentence.

The challenge, then, is to define the meaning of the words in the sentence so that its meaning can be derived in a fully compositional fashion. As we know from Chap. 7, the intransitive verb “paints” refers to the well-defined sets of objects in the model that satisfy the property of painting. We can analyze the meaning of the common noun “dancer” in a similar fashion, that is, as the set of objects that are dancers in the model. The expression that requires the most attention, and is most crucial to providing a compositional account of the meaning of the whole sentence, is the determiner “every”. Clearly, “every” is the element in the sentence that expresses the relation between the two sets referred to by, respectively, “dancer” and “paints”. In our terms, “every” performs a function similar to that of transitive verbs—it expresses a relation between two elements of the model. The only difference is that, whereas transitive verbs express relations between individual objects, “every” expresses a relation between sets. We can express this relation in terms of the following function: The function that takes an ordered pair of two sets,  $X$  and  $Y$ , and delivers the truth-value 1 if, and only if,  $X$  is a subset of  $Y$ , 0 otherwise.

(2)  $\langle X, Y \rangle \rightarrow 1$  if  $X$  is a subset of  $Y$ , 0 otherwise

The symbolization in (2) uses capital letters to make clear that the input of the function,  $\langle X, Y \rangle$ , is an ordered pair composed of sets—such as the sets of dancers and the set of painters—and not of individual objects—such as Alexandra and Pablo.

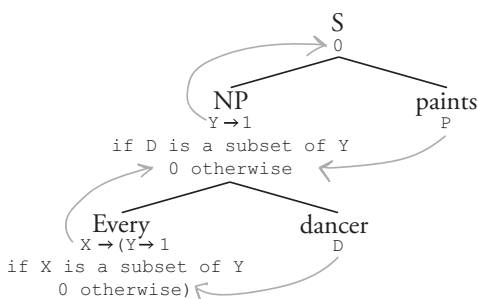
The function in (2) correctly captures the relation expressed by “every”, which is a relation between two sets  $X$  and  $Y$  that is true just in case  $X$  is a subset of  $Y$ . Applied to the pair comprising the set of dancers and the set of painters, it delivers the truth-value 1 if, and only if, the dancers are a subset of the painters. The function, however, does not allow us to derive the meaning of the sentence compositionally, in the order that is dictated by the grammatical structure of the sentence. The structure of

the sentence, in fact, tells us that first the meaning of “every” combines with the meaning of “dancer” and, then, the meaning of “every dancer” combines with meaning of “paints” to deliver the meaning of the whole sentence. To achieve this goal we must apply the same strategy we applied to transitive verbs in the previous chapter. We must, in other words, apply the currying transformation to (2) so that its input arguments can be applied in the order that is dictated by the grammatical structure of the sentence. The function in (3) is what we need.

$$(3) X \rightarrow (Y \rightarrow 1 \text{ if } X \text{ is a subset of } Y, 0 \text{ otherwise})$$

The function in (3) is a function that takes a first set  $X$  as its input to deliver yet another function that takes a second set  $Y$  as its input in order to deliver the value 1 if, and only if,  $X$  a subset of  $Y$ . This function is fully equivalent to (2) as it delivers the same output value under the same circumstances. It differs from (2) only in that the two sets  $X$  and  $Y$  are fed to the function one at the time, rather than together as a single ordered pair.

Thanks to (3) we can now derive the meaning of the sentence “every dancer paints” in a compositional fashion. Consider, once again, the grammatical structure of the sentence together with the meanings of the simple expressions occurring in it. This is exemplified in Fig. 9.6 where, for practicality, we have symbolized the set of dancers with the capital letter  $D$  and the set of painters with the capital letter  $P$ . The function referred to by the determiner “every” is the function we defined in (3) and it demands a set as its input argument. The noun “dancer” provides just that. If we, then, apply the set  $D$  to the function, we obtain, as its output, the function that maps  $Y$  into the truth-value 1 if  $D$  is a subset of  $Y$ , to 0 otherwise. This function is the meaning of the noun phrase (NP) “every dancer”. It is a function that demands another set as its input. The intransitive verb “paints” provides just that. By applying



**Fig. 9.6** A fully compositional derivation of the meaning of the sentence “every dancer paints”. In the first step, the set of dancers  $D$  is applied to the function  $X \rightarrow (Y \rightarrow 1 \text{ if } X \text{ is a subset of } Y, 0 \text{ otherwise})$ , which is the reference of the determiner “every”. This application delivers the function  $Y \rightarrow 1 \text{ if } D \text{ is a subset of } Y, 0 \text{ otherwise}$ , which is the reference of the NP “every dancer”. In the second step, the set of painters  $P$  is applied to the function  $Y \rightarrow 1 \text{ if } D \text{ is a subset of } Y, 0 \text{ otherwise}$ . This application delivers, in the model of interpretation we are adopting, the truth-value 0, which is the reference of the sentence

the set  $P$  to the function, we finally obtain the truth-value of the sentence, which, in the model at hand, is 0, as it is not the case that  $D$  is a subset of  $P$ .

Once again, by relying on the formal properties of our apparatus of objects, sets, and functions, we have been able to provide a step-by-step derivation of the meaning of the sentence, based on the meanings of the individual words in the sentence and their grammatical arrangement.

There is something interesting to observe about the compositional process we have exemplified in picture 9.6. In Chap. 7, we saw that the meaning of a simple declarative sentence such as “Kazimir paints” is derived compositionally by applying the reference of the subject “Kazimir” as the input of the function referred to by the verb “paints”. In the case of a sentence such as “every dancer paints” the process is inverted. In this case, it is the reference of the verb “paints” that applies as the input of the function referred to by the subject “every dancer” in order to deliver the meaning of the whole sentence. This observation gives us yet another glimpse of how the mathematical structure at the basis of Montague’s theory can be exploited to construct functions of higher and higher complexity to derive the meaning of grammatical structures of higher and higher complexity.

It is easy to extend the analysis we discussed for “every” to the other determiners. We can describe the reference of the determiner “no” with the function in (4).

(4)  $X \rightarrow (Y \rightarrow 1$  if the intersection between  $X$  and  $Y$  is empty, 0 otherwise)

This is a function that delivers the truth-value 1 if, and only if, the two input sets have an empty intersection, that is, if there is no entity in the model that belongs to both sets. Similarly, we can describe the reference of “some” as the function in (5).

(5)  $X \rightarrow (Y \rightarrow 1$  if the intersection between  $X$  and  $Y$  is non-empty, 0 otherwise)

This function delivers the truth-value 1 if, and only if, the intersection between the two input sets is not empty, that is, if there is at least one object in the model that belongs to both sets.

## References and Remarks

The study of quantification has played a central role in logic since its inception. In his *Prior Analytics*, Aristotle provides a first systematic analysis of the quantifiers “all”, “no”, “some”, and “not all” on the basis of how they contribute to different logical syllogisms. The modern study of quantifiers must be credited, once again, to Frege. He first formulated the language of predicate logic and, with it, the modern notions of universal and existential quantifiers, commonly notated in today’s predicate calculus as  $\forall$  and  $\exists$  (Frege 1879, 1893/1903). To Frege, we also owe the idea that quantifiers represent relations between properties—also called second-order relations. The modern formulation of this idea as a model-theory for the interpretation of quantifiers in a formal language is due to Mostowski (Mostowski 1957). It is known as the theory of *generalized quantifiers* because it offers a framework for the

interpretation of quantifiers that can be easily generalized to all quantifiers, beyond the universal and existential ones. The use of the theory of generalized quantifiers as a framework for the interpretation of noun phrases in natural language can already be found in Montague (Montague 1973), although a more thorough discussion of the notion of general quantification in the context of natural language would only take place a few years later thanks to the contributions of scholars such as John Barwise and Robin Cooper (Barwise and Cooper 1981), James Higginbotham and Robert May (Higginbotham and May 1981), and Edward L. Keenan & Jonathan Stavi (Keenan and Stavi 1986). To those interested in an overview of the state of the art in the study of quantification in formal and natural language, we recommend the entry “Generalized quantifiers in linguistics and logic” by Keenan and Westerståhl in the *Handbook of Logic and Language* (Keenan and Westerståhl 2011).

A somewhat surprising consequence of the theory of generalized quantifiers—particularly, when contrasted with the more traditional predicate calculus of Frege—is the parallel between determiners and transitive verbs. Despite being different grammatical categories, they are regarded as performing similar semantic functions. In fact, they both express transitive relations—between properties in the case of determiners and between individuals in the case of transitive verbs. This at first counterintuitive parallel is discussed in detail in Chap. 8 of Richard Larson and Gabriel Segal’s *Knowledge of Meaning* (Larson and Segal 1995).

A well-known limitation of the theory of generalized quantifiers in the context of natural language noun phrases is that it cannot be applied straightforwardly to noun phrases in positions different from that of grammatical subject. Consider, as an example, the sentence “Kazimir admires every painter”. According to the theory of generalized quantifiers, the NP “every painter” expresses a function from sets into truth-values (more exactly, the function that maps a set  $X$  into the truth-value 1 if, and only if, the set of painters is a subset of  $X$ ). This function is incompatible with the function referred to by the transitive verb “admires”, which, as we saw in Chap. 7, is a function from individuals into a function from individuals into truth-values. The problem is, therefore, that the two functions cannot combine with one another on the basis of the mechanism of functional application. On the one hand, the reference of “admires” cannot be applied as the input of the function expressed by “every painter”—this function requires a set as its input whereas “admires” provides a relation. On the other hand, the reference of “every painter” cannot be applied as the input of the function expressed by “admire”—this function requires an individual object as its input whereas “every painter” provides a function from sets to truth-values. One solution to this problem is to modify the relation expressed by transitive verbs so to accommodate the semantic type of noun phrases in object position. This is the solution adopted by Montague himself in his article “The proper treatment of quantification in ordinary English” (Montague 1973). Another solution is to assign a different semantic type to noun phrases occurring in object position—a semantic type specifically designed to allow them to combine compositionally to transitive verbs. This solution can be implemented systematically through the use of general rules of *type-shifting*—rules, that is, that can change the semantic type of an expression whenever necessary to solve a combinatorial clash. This type of solution is

discussed in Partee and Root's article "Generalized Conjunction and Type Ambiguity" (Partee and Rooth 1983; see also Partee 1986) and in Herman Hendriks's "Type Change in Semantics: The Scope of Quantification and Coordination" (Hendriks 1987). Both these solutions deal with the mismatch *in situ*—that is, in the place in the structure where the noun phrase is originally merged. An altogether different type of approach, initiated with Robert May's PhD dissertation "The grammar of Quantification" (May 1977; see also May 1985), addresses the problem by applying a syntactic transformation to the grammatical structure of the sentence before it is interpreted compositionally. This transformation is known as *Quantifier Raising* because it rearranges sentential structures by re-merging the noun phrases embedding it at the top of the sentential node. Application of this transformation allows a compositional analysis of quantifiers that assigns them a consistent semantic type. Both approaches have their advantages and disadvantages. We refer the readers interested in this discussion to Eddy G. Ruys and Yoad Winter's entry "Quantifier scope in formal linguistics" in the *Handbook of Philosophical Logic* (Ruys and Winter 2011). A common goal of all these different approaches, besides that of solving the compositionality challenge posed by non-subject noun phrases, is that of capturing the scopal ambiguities that we see in sentences with multiple quantifiers. As an example, consider the sentence "every painter admires a dancer". As Frege himself had already noticed (Frege 1892), this sentence is ambiguous as it is passible of two different interpretations. On the one hand, it can mean that for every painter there is a possibly different dancer whom the painter admires. On the other, it can mean that there is a single dancer whom every painter admires. One of the main goals of theories of quantification in natural language is that of providing systematic accounts of this ambiguities. Ruys and Winter's article, mentioned above, offers a survey of how different theories of natural language quantification deal with these issues.

## References

- Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4, 159–219.
- Frege, G. (1879). *Begriffsschrift. Eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle: Verlag von Louis Nebert.
- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100(1), 25–50.
- Frege, G. (1893/1903). *Grundgesetze der Arithmetik*. Jena: Verlag Hermann Pohle.
- Hendriks, H. (1987). Type change in semantics: The scope of quantification and coordination. In E. Klein & J. van Benthem (Eds.), *Categories, polymorphism and unification* (pp. 96–119). Edinburgh: Centre for Cognitive Science.
- Higginbotham, J., & May, R. (1981). Questions, quantifiers and crossing. *Linguistic Review*, 1, 41–79.
- Keenan, E. L., & Stavi, J. (1986). A semantic characterization of natural language determiners. *Linguistics and Philosophy*, 9, 253–326.

- Keenan, E. L., & Westerståhl, D. (2011). Generalized quantifiers in linguistics and logic (revised version). In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (2nd ed., pp. 859–910). Amsterdam: Elsevier.
- Larson, R., & Segal, G. (1995). *Knowledge of meaning*. Cambridge, MA: MIT Press.
- May, R. (1977). The grammar of quantification. Doctoral dissertation, MIT.
- May, R. (1985). *Logical form: Its structure and derivation*. Cambridge, MA: MIT Press.
- Montague, R. (1973). The proper treatment of quantification in ordinary English. In *Approaches to natural language* (pp. 221–242). Dordrecht: D. Reidel.
- Mostowski, A. (1957). On a generalization of quantifiers. *Fundamenta Mathematicae*, 44(1), 12–36.
- Partee, B. (1986). Noun phrase interpretation and type-shifting principles. In J. Groenendijk, D. de Jongh, & M. Stockhof (Eds.), *Studies in discourse representation theory and the theory of generalized quantifiers* (pp. 115–143). Dordrecht: Foris.
- Partee, B., & Rooth, M. (1983). Generalized conjunction and type ambiguity. In R. Bauerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, use, and interpretation of language* (pp. 361–383). Berlin: de Gruyter.
- Ruys, E. G., & Winter, Y. (2011). Quantifier scope in formal linguistics. In *Handbook of philosophical logic* (pp. 159–225). Dordrecht: Springer.

# Chapter 10

## Meaning and Description



### Definite Descriptions

In this chapter, we will briefly focus on a determiner that has received special attention from philosophers, logicians, and linguists and that will play a central role throughout the rest of the book—the *definite article* “the”. The definite article combines with a common noun to produce a so-called *definite description*, which is a noun-phrase of the type exemplified by “the painter” in (1).

(1) The painter sings

The determiner “the” can be analyzed, on a parallel with other determiners, as a relation between two properties (and their corresponding sets), although one that comes with an extra restriction. A canonical way of expressing its meaning is given in (2).

(2)  $X \rightarrow Y \rightarrow 1$  if  $X$  contains exactly one entity and  $X$  is a subset of  $Y$ , 0 otherwise

According to (2), the relation expressed by “the” overlaps partially with that expressed by “every”: It is a relation between two sets  $X$  and  $Y$  which is true if, and only if,  $X$  is a subset of  $Y$ . However, “the” further requires that set  $X$  contains exactly one member. Sentence (1), according to this analysis, is true if, and only if, the model in which the sentence is evaluated includes exactly one painter and that painter also sings.

This analysis captures the intuition that definite descriptions of the form *the + common noun* are interpretable only in circumstances in which the common noun singles out a unique object. In effect, if we were asked, “can you pass me the book?” in a context in which there is more than one book, we would not be able to interpret the definite description “the book” and would probably have to ask which book the speaker exactly meant.



## Reference and Description

Definite descriptions are somewhat similar to proper names in that they are means of reference to individual objects in a model. Amongst English speakers, the definite description “the Russian painter author of the Suprematist manifesto”, for example, performs the same function as the proper name “Kazimir Malevich” in singling out the same individual and can, in fact, be used as a synonym.

This similarity, however, is only apparent as proper names and definite descriptions perform their function in substantially different ways. The reference of a proper name is stipulated in the lexicon of the language. In order to know what the meaning of the proper name “Kazimir Malevich” is, we need to know what its reference is. The referent of a description, on the other hand, is determined indirectly as the object that uniquely satisfies a certain property. In fact, whereas proper names and definite descriptions may perform similar functions, they do not have, strictly speaking, the same meaning. The meaning of a proper name is its *reference*—the object it refers to. The meaning of a definite description is a *description*—that is, a property that univocally individuates an object. Whereas the meaning of a proper name is *what* the name refers to, the meaning of a definite description is *how* the referent is referred to.

We can appreciate this difference more concretely by observing that the meaning of a definite description is free to change depending on the circumstances, whereas the meaning of a proper name is fixed and constant. Suppose Kazimir is standing in front of you. You may refer to him either by using the proper name “Kazimir” or the description “the person standing in front of me”. Both the proper name and the definite description do the job of singling out the individual Kazimir. Suppose, however, that Kazimir now moves away and Frida becomes the person standing in front of you. The meaning of the definite description “the person standing in front of me” now adapts to the changed circumstances and ends up singling out Frida. The proper name “Kazimir”, on the contrary, continues to refer to Kazimir. Proper names denote their referent directly, like labels we put on objects. Definite descriptions point at their referent indirectly, through the mediation of a description—a property that has the ability to uniquely identify an individual in the circumstances at hand. Whereas names tell us *what* we are talking about, definite descriptions tell us *how* we are talking about it. Whereas proper names point at things, definite descriptions describe them.

## References and Remarks

Definite descriptions have attracted the attention of philosophers, logicians, and linguists for a number of different reasons. In the context of our discussion, there are two issues that must be mentioned. The first concerns their logical type. In our presentation, we have treated the definite article as a generalized quantifier, expressing

the same type of relation expressed by the quantifiers discussed in Chap. 9—that is, a relation between sets. This type of analysis was originally formulated by Bertrand Russell in his article “On Denoting” (Russell 1905) and is commonly formalized by making use of the *iota*-operator from his and Alfred North Whitehead’s *Principia Mathematica* (Whitehead and Russell 1910). In short, the *iota*-operator expresses a relation between two properties *A* and *B* which is true if, and only if, there is a unique individual that satisfies *A* and that individual also satisfies *B*. Russell’s analysis is also the analysis maintained by Montague in his treatment of English (Montague 1973). A different approach, which can be traced back to Frege (Frege 1892), assigns to definite descriptions the same semantic type as proper names. According to this analysis, the definite article “the” expresses a function that takes a property as its input and delivers an individual object as its output, the individual object that uniquely satisfies the input property. When applied to the property referred to by “painter”—as in the definite description “the painter”—“the” delivers the unique individual in the model who satisfies the property of being a painter. This analysis reduces significantly the difference between proper names and definite descriptions, which are assigned references of the same semantic type. This is the analysis adopted by Heim and Kratzer in their textbook (Heim and Kratzer 1998). A canonical reference to fully appreciate the philosophical implications of this difference in analysis is Neale (1990), which also amounts to a sophisticated defense of Russell’s view of descriptions.

A different—yet related—point of discussion in the semantics of definite descriptions concerns whether their existential commitment is asserted or presupposed. This discussion has pivoted around sentences such as “the king of France is bald”. As we all know, France is a republic and, therefore, has no king. Hence, the question is whether the sentence is just plainly false—it is not the case that the king of France is bald—or, rather, meaningless—that is, it cannot be assigned a truth-value because the conditions for evaluating its truth or falsity are simply not met. Russell’s analysis and, consequently, the one we have endorsed above, predicts that the sentence is false. According to this analysis, “the” requires that there is exactly one object at the intersection between its two input sets. If there is no such object, the condition is not met and the function delivers the value 0. This view, defended by Russell, was contested by Peter Frederick Strawson in his “On Referring” (Strawson 1950). There, Strawson advances the view that the existence of a unique object that satisfies the description is a precondition—or, to use a more technical term, a *presupposition*—that must be met for the definite description to refer. If this presupposition is not met, the description fails to refer and, henceforth, to contribute a meaning. According to Strawson, then, the sentence “the king of France is bald” is not false but meaningless.

Finally, it is worth mentioning yet another, more recent class of approaches to the semantics of definite descriptions that were developed in parallel by Hans Kamp (Kamp 1981) and Irene Heim (Heim 1982). Kamp’s and Heim’s approaches share with Strawson the view that definite descriptions are referential, rather than quantificational, but also add a formalization of the relation they hold to the context in which they occur. These approaches set the foundations for a type of semantic

theorizing known as *Discourse Representation Theory*, which is a view that, instead of analyzing the meaning of sentences in isolation, looks at them as parts of larger linguistic structures, called discourses. The canonical reference for the readers interested in this type of approach is Hans Kamp and Uwe Reyle's book *From Discourse to Logic* (Kamp and Reyle 1993).

## References

- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100(1), 25–50.
- Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. PhD thesis. Amherst: University of Massachusetts.
- Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Oxford: Blackwell.
- Kamp, H. (1981). A theory of truth and semantic representation. In J. A. G. Groenendijk, T. M. V. Janssen, & M. B. J. Stokhof (Eds.), *Formal methods in the study of language* (Mathematical centre tracts) (Vol. 135, pp. 277–322). Amsterdam: Mathematisch Centrum.
- Kamp, H., & Reyle, U. (1993). *From discourse to logic*. Dordrecht: Kluwer.
- Montague, R. (1973). The proper treatment of quantification in ordinary English. In *Approaches to natural language* (pp. 221–242). Dordrecht: D. Reidel.
- Neale, S. (1990). *Descriptions*. Cambridge, MA: MIT Press.
- Russell, B. (1905). On denoting. *Mind*, 14(56), 479–493.
- Strawson, P. F. (1950). On referring. *Mind*, 59(235), 320–344.
- Whitehead, A. N., & Russell, B. (1910). *Principia mathematica* (Vol. 3). Cambridge: Cambridge University Press.

# Chapter 11

## Meaning and Possibility



In the previous chapters, we familiarized ourselves with the core workings of Montague's approach to natural language meaning and, in particular, with how the mathematical structure at its basis allowed him to express the compositional nature of meaning. Compositionality, however, is not the only aspect of natural language meaning Montague wanted to capture with his theory. Another fundamental goal of his theory is that of expressing the logical properties of natural language expressions—the fact, that is, that natural language expressions contribute together to a net of deductive relations of different sorts. To achieve this goal, Montague chose a special type of model-theory, which draws a distinction between two different classes of objects within the model: individual objects, of the type we considered so far, and *possible worlds*, a special class of entities which will be the main topic of discussion in this and the following two chapters. The adoption of this type of model will have important implications for our inquiry into linguistic meaning. We will return to these implications in Chap. 14.

### Truth-Values

On a parallel with the formal language we discussed in Chap. 6, we have so far maintained that declarative sentences refer to *truth-values*. A declarative sentence refers to the truth-value 1, if it is true in the model, to the truth-value 0, if it is false. This is a reasonable assumption to be made in the context of a formal language. The goal of a formal language, as we saw, is that of describing a model in a precise and unambiguous way. A formal language should, henceforth, be able to tell us precisely and unambiguously whether a certain condition is met or not by the model it describes. But does the same hold for natural language?

Some readers may find it already awkward that truth and falsity are conceived in our current theory as objects we can refer to by means of linguistic expressions, in

the same way as we refer to objects by means of proper names or to sets by means of intransitive verbs. Let us not forget, however, that our current understanding of the notion of object in a model is a very abstract one. As we saw in Chap. 6, all sorts of things can be objects in a model, as long as they can be subject to the basic operations of set-theory. A model can include prices as well as beliefs, chairs as well as molecules, dreams as well as truth-values.

There are, nonetheless, at least three reasons to doubt that natural language declarative sentences refer to truth-values. To begin with, if knowing the meaning of a sentence is equivalent to knowing its truth-value, then, knowing the meaning of a sentence is equivalent to knowing whether the sentence is true or false. This seems incorrect. Consider the following sentence:

- (1) Kazimir Malevich was born in Ukraine to Polish parents

Those unfamiliar with the details of Kazimir Malevich's life may not be able to tell whether (1) is true or false and, consequently, may not be able to identify the truth-value of the sentence. Should we conclude that these readers are failing to understand the meaning of (1)? Certainly not. Not knowing whether (1) is true or false does not prevent us from having a perfect understanding of what (1) means.

The second problem is that, if the meaning of a declarative sentence is its truth-value, then all true sentences have the same meaning. This is because all true sentences refer to the same object, that is, the truth-value 1. Since, according to our hypothesis, the meaning of an expression is the object it refers to in the model, all expressions that refer to the same object also have the same meaning. Hence, since all true sentences refer to the same object—the truth-value 1—we must conclude that they also have the same meaning. Of course, the same goes for all false sentences. Since they all refer to the truth-value 0, we must conclude that they all have the same meaning. This is also incorrect. To see why, compare (1) with (2).

- (2) Pablo Picasso was born in Spain to Spanish parents

As it happens, it is true that Kazimir Malevich was born in Ukraine to Polish parents. Hence, (1) refers to the truth-value 1. However, it is also true that Pablo Picasso was born in Spain to Spanish parents. Hence, also sentence (2) refers to the truth-value 1. Should we conclude that, since (1) and (2) refer to the same truth-value, they also have the same meaning? Of course not. While (1) and (2) happen to be both true, they clearly describe different states of affairs and, in fact, contribute different information.

The third problem with the idea that the reference of a sentence is its truth value is that it fails to capture the logical relations that hold between declarative sentences. Consider the following three sentences:

- (3) Kazimir paints  
 (4) Somebody paints  
 (5) Nobody paints

Suppose we are asked to judge the truth of these sentences but the only information we are given is that (3) is true. This is how we can reason. If it is true that

Kazimir paints, then there is certainly at least one person who paints. Hence, if (3) is true, we can infer, as a matter of logical deduction, that (4) is true as well. Furthermore, if it is the case that Kazimir paints, then there is at least one person who paints and that is all the evidence we need to conclude that (5) is false. We find, to summarize, that (3) *entails* (4)—if (3) is true, then (4) is as well—and that (3) and (5) are *contrary* sentences—if (3) is true, then (5) is false. The problem for our current theory is that it does not provide an account of these logical relations. All our theory does is assign truth-values to sentences. In this way, however, it tells us nothing about the logical relations that hold between the truth-values of different sentences. This is an especially negative result. The ability to identify logical relations between sentences is a central part of the speakers' ability to use language meaningfully. Speakers use natural language to reason, argue, formulate assertions, consider hypotheses, prove, and express contrary and contradictory sentences. A satisfactory theory of linguistic meaning must be able to account for this capacity.

## Truth-Conditions

Our current theory identifies the meaning of declarative sentences with their truth-value. However, our intuition of speakers tells us that to know the meaning of a declarative sentence is not to know whether the sentence is true or false. To understand a sentence is, rather, to understand what would make it true—that is, what *conditions* the model must meet for the sentence to be true.

When we read sentence (1), we understand that, in order for the sentence to be true, it must be the case that the individual entity Kazimir Malevich belongs to the set of those entities that were born in Ukraine to Polish parents. The sentence itself does not tell us that this is the case. Rather, it tells us what has to be the case for it to be true.

Similarly, we find that (1) and (2) have different meanings because they are true under different conditions. In the current circumstances, it happens that both sentences are true. However, this is not necessarily so. Things could have been different and one sentence could have turned out false, while the other true. For example, Kazimir Malevich might have been born in, say, Poland and Pablo Picasso in Italy. The two sentences inform us of different conditions that the model must meet for them to be true.

Finally, it is our knowledge of the conditions that make different sentences true that allows us to appreciate the logical relations between them. We are led to conclude that (3) entails (4) because we find that the conditions that make (3) true make (4) true as well. As we saw, in order for (3) to be true, it must be the case that Kazimir Malevich—a person—belongs to the set of painters. These conditions are sufficient to make also (4) true. In Chap. 9, we saw that a sentence such as (4) is true if, and only if, the intersection between those entities that are persons and those entities that paint is non-empty. Hence, any model where (3) is true is also automatically a model where (4) is true. Similarly, we cannot think of a model where (3) and (5)

are simultaneously true (this is what makes them contrary sentences; notice, incidentally, that they could be both false, that is, they are not contradictory sentences). In order for (5) to be true, it must be the case that the set of those who paint has an empty intersection with the set of people. But this configuration is incompatible with the one that makes (3) true (that is, the one whereby Kazimir is a person and he paints). This is why we conclude that (3) and (5) are contrary sentences.

The lesson we learn is that understanding the meaning of a sentence is not the same as identifying its truth-value. Rather, it is identifying its *truth-conditions*. But what are truth-conditions exactly? And how can we construct such a notion in our model-theory?

## Possible Worlds

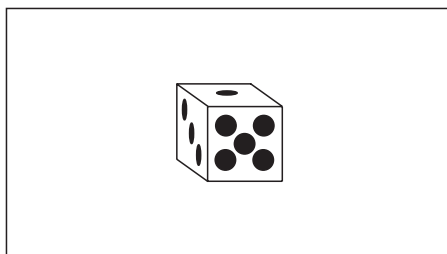
There are different ways to express the truth-conditions of a sentence in a model. The framework we are about to adopt is known as *possible world semantics*. This framework was developed from ideas put forward by the mathematician Rudolf Carnap and the philosopher and logician Saul Kripke, although its core insights can already be found in the writings of the philosopher Gottfried Wilhelm Leibniz.

Let us consider for a moment a rather peculiar type of model, one whose domain includes a single object: a die (Fig. 11.1). A model with one, single object is indeed a peculiar type of model; nonetheless, a legitimate one. As we saw, all a model does is providing a quantitative description of a set of objects and their properties. In principle, nothing stops us from constructing a model that describes a single object.

Typically, a die is used whenever one needs a random natural number between 1 and 6, for example, when playing a board game. The die is thrown on the table and the number that shows up on its upward looking face is the one that is usually taken into consideration. Mathematicians have developed precise methods to determine the *probability* that a certain number (or set of numbers) will be produced by a die. For example, given that there are six possible numbers and only three of them are even, the probability of obtaining an even number when rolling the die is calculated as follows:

$$(6) \quad 3/6 = 0.5$$

**Fig. 11.1** A model including a single object, a die



The formula in (6) tells us that there is a 0.5—that is, 50%—probability that the die will produce an even number. The formula clearly expresses a relation between quantities. But quantities of what? What is the model underlying the reasoning described by (6)? The formula tells us that the chances of obtaining an even number when rolling the die are obtained by dividing three quantities of something by six quantities of something else. What are these quantities quantities of?

The question is particularly puzzling if we consider that the model we are dealing with contains a single object—the die. A model with only one object is not of much help: How can we reason about three quantities of this and six quantities of that when all we have is one, single object? To address this problem and provide a satisfactory model of interpretation of formulas such as (6), we must enrich the model with a new dimension—the dimension of *possibility*. What we need, in a nutshell, is a model that does not simply provide a description of a state of affairs *as it is*, but also *as it could be*.

Suppose you are about to roll the die. At this stage, you cannot predict in which position the die will stop, hence, you cannot predict which number it will deliver. Yet, the possibilities are not limitless. When it comes to what number will show up on the upward looking face of the die, there are six possible outcomes:

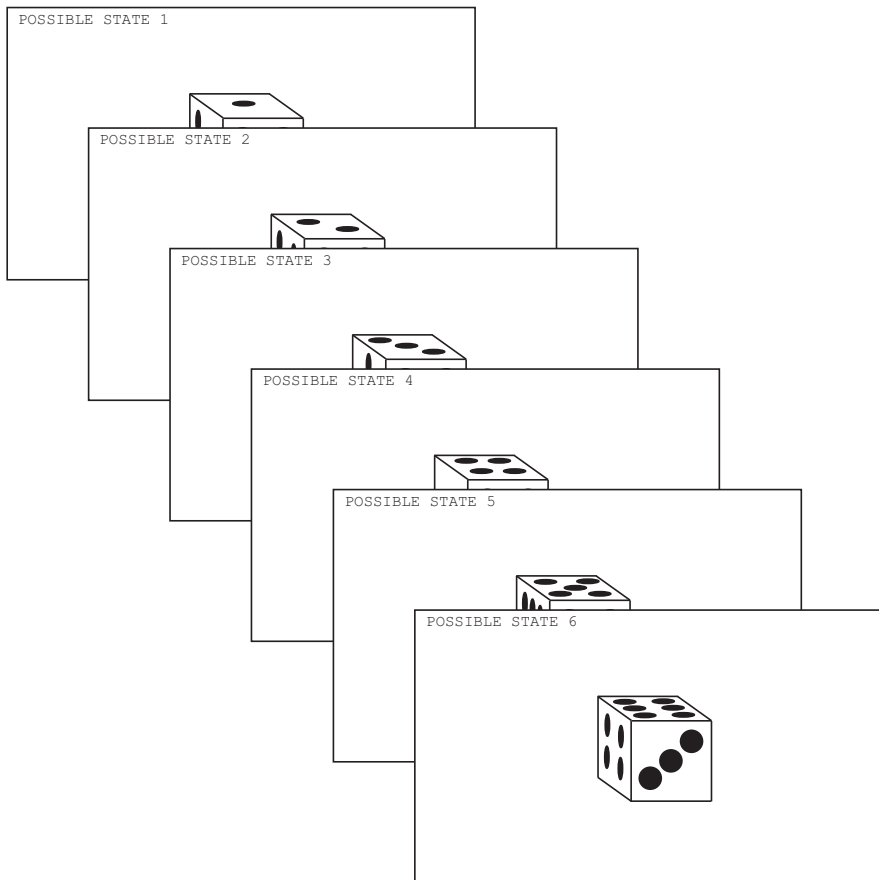
- Possible outcome 1: The die shows the number 1 on its upward looking face
- Possible outcome 2: The die shows the number 2 on its upward looking face
- Possible outcome 3: The die shows the number 3 on its upward looking face
- Possible outcome 4: The die shows the number 4 on its upward looking face
- Possible outcome 5: The die shows the number 5 on its upward looking face
- Possible outcome 6: The die shows the number 6 on its upward looking face

The calculation in (6) expresses a relation between these six possible outcomes. Given that of these six possible outcomes, only three are characterized by the die delivering an even number, there are three possible outcomes out of six— $3/6$ , hence, a 50% chance—to obtain an even number when rolling the die.

The model of interpretation of the formula in (6) is, henceforth, a model that includes an object—the die—along with the *possible states* in which this object may find itself. It is these possible states that are the objects of quantification of the formula in (6).

Possible states of affairs are commonly referred to by logicians as *possible worlds* (or *possible universes*). So far, we have assumed that the model of interpretation of a language is the totality of the objects, properties, and relations that are relevant to its interpretation. As we saw, logicians refer to it as the *world* (or *universe*) of interpretation of the language. The model we are now considering is richer than that, since it does not only include all the *actual* objects, properties, and relations that are relevant to the interpretation of (6) but, in fact, also the *possible* ones. The model of interpretation of (6) is made of not one but six possible worlds (or possible universes), which, together, describe the totality of the possible objects, properties, and relations that are relevant to its interpretation. We can visualize such a model as a stack of possible descriptions of the same die, each representing a different possible state of the die, as in Fig. 11.2.





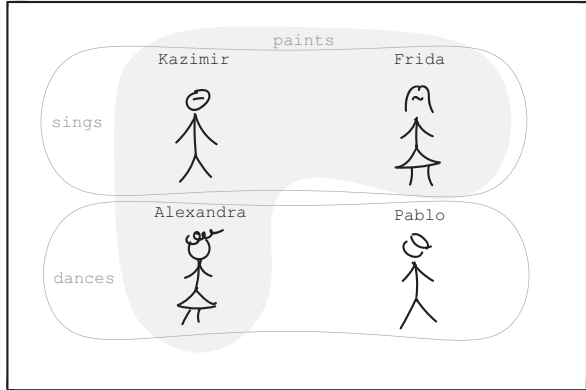
**Fig. 11.2** A model including a die and six possible states of the die, each characterized by a different number on the upward looking face of the die

Of course, we can apply the concept of possible world also to our model—the model we have used in the previous chapters to illustrate Montague’s theory of natural language meaning. Consider the model as it is reproduced in Fig. 11.3. In this model, we find that Kazimir, Frida, and Alexandra paint, Alexandra and Pablo dance, and Kazimir and Frida sing.

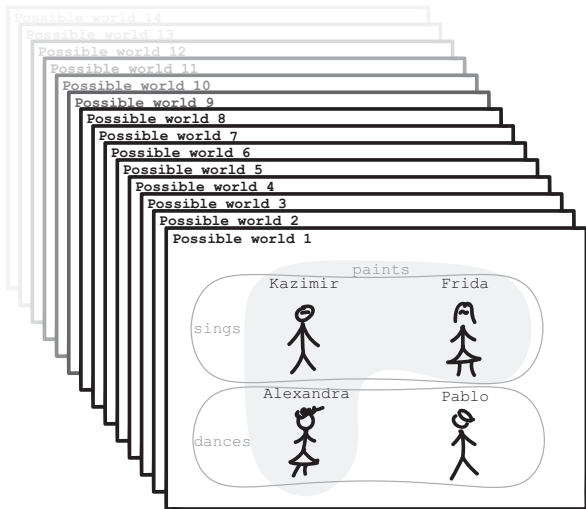
This, however, is only one of many possible scenarios. Things may have gone differently. For example, Frida may not have been painting, but dancing, and Pablo may not have been dancing, but singing. In fact, all combinations are possible for all objects, properties, and relations in the model. As we did for the die, we can represent the totality of these possible arrangements as a set of possible worlds, as in Fig. 11.4.

The model we obtain, enriched with the dimension of possibility, now represents not one, but all possible arrangements of its objects, properties, and relations.

**Fig. 11.3** The model of interpretation from Chap. 7, representing the four individuals Kazimir, Frida, Pablo, and Alexandra and the three properties of painting, singing, and dancing



**Fig. 11.4** Our model of interpretation reconstructed as a stack of possible worlds, each providing a different arrangement of the objects, properties, and relations in the original model



Possible worlds, as we will see in the following section, allow us to construct the notion of truth-conditions within the logical structure of our model.

## Propositions

As we saw above, the truth-conditions of a declarative sentence are the conditions that a state of affairs must meet for the sentence to be true. Consider, as an example, sentence (3), repeated below:

- (3) Kazimir paints

































The truth-conditions of (3) are, again, the conditions that must be met by the circumstances of evaluation of (3) for the sentence to be true. In the case of (3), these conditions dictate that, for the sentence to be true, it must be the case that the individual Kazimir belongs to the set of objects who paint. The truth-conditions of a declarative sentence are, henceforth, a *property of states of affairs*—a property a state of affairs must meet to make the sentence true.

The notion of a property is not new to us. We have already seen that intransitive verbs such as “paints” and “sings” and proper nouns such as “painter” and “singer” refer to properties. These properties are *properties of objects*, such as Kazimir or Frida. In fact, we can characterize them in terms of corresponding sets of objects. The property denoted by “paint” is the set of all objects that paint. The property characterized by “singer” is the set of all objects that are singers. The same notion can be generalized to other types of objects, such as possible worlds. In fact, if a property of objects corresponds to the set of objects that satisfy that property, a property of state of affairs can be similarly identified with the set of possible worlds that satisfy that property. The property of being a state of affairs where Kazimir paints, for example, is equivalent to the set of possible worlds where Kazimir belongs to the set of those who paint.

We see that, in a framework that encompasses possible worlds amongst its objects of quantification, we can characterize truth-conditions as *well-defined sets of possible worlds*, that is, sets of possible worlds that are identified by a common property. We will refer to these sets of possible worlds as *propositions* and we will take propositions to be the referents of declarative sentences. Accordingly, we will say that sentence (3) refers to the set of all possible worlds that share the property that, in them, Kazimir belongs to the set of those who paint. We will refer to this set of possible worlds as the proposition referred to by (3).

If we think of a model as we have done in this chapter—that is, as describing a space of logical possibilities—we can think of a declarative sentence as performing the function of dividing the possibility space into two regions: the region of possibilities that make the sentence true and the region of possibilities that make the sentence false. We can, for example, think of sentence (3) as performing the function of distinguishing the possible worlds in the model where Kazimir belongs to the set of those who paint from those where he does not. More concretely, if we restrict our attention to the four individuals in our model and the property of painting, we find that there are sixteen possible arrangements, which we can represent schematically as in Table 11.1. Each row in the table represents a possible world. Each column represents an individual. The brush sign in a cell indicates that the individual of the corresponding column belongs to the set of those who paint in the possible world of the corresponding row. World 1, for example, is a world where everybody paints. World 16 is a world where nobody paints. World 4 is a world where Kazimir and Frida paint but Alexandra and Pablo do not. And so on. There are eight worlds (worlds 1–8) where Kazimir belongs to the set of those who paint. These possible worlds are the set of possible worlds referred to by sentence (3), because they are all the possible worlds that make sentence (3) true. This set of possible worlds is the proposition referred to by (3).

**Table 11.1** The sixteen possible arrangement of the four individuals Kazimir, Frida, Alexandra, and Pablo and the property of painting

	Kazimir	Frida	Alexandra	Pablo
Possible world 1				
Possible world 2				
Possible world 3				
Possible world 4				
Possible world 5				
Possible world 6				
Possible world 7				
Possible world 8				
Possible world 9				
Possible world 10				
Possible world 11				
Possible world 12				
Possible world 13				
Possible world 14				
Possible world 15				
Possible world 16				

Each row in the table represents a possible arrangement, hence a possible world. Each column represents an individual. The brush sign in a cell indicates that the individual of the corresponding column belongs to the set of those who paint in the possible world of the corresponding row

## The Algebra of Entailment

In the previous section we have seen that declarative sentences denote propositions and that propositions correspond to sets of possible worlds. It is time we see how this framework allows us to address the problems raised at the beginning of the chapter.

The first problem we considered was that understanding a declarative sentence does not correspond to understanding whether the sentence is true or false. According to our new framework, to understand the meaning of a sentence now corresponds to understanding its corresponding proposition—that is, what property the model must satisfy for the sentence to be true. When, for example, we understand (1), repeated below, we understand that, for (1) to be true, it has to be the case that Kazimir Malevich belongs to the set of those who were born in Ukraine to Polish parents.

- (1) Kazimir Malevich was born in Ukraine to Polish parents

The second problem we considered at the beginning of the chapter was that sentences with equivalent truth-values do not have equivalent meanings. In the new framework of possible worlds and propositions, this is no longer a problem. Sentences (1) and (2), also repeated below, are indeed both true.

## (2) Pablo Picasso was born in Spain to Spanish parents

However, they refer to different propositions. Sentence (1) refers to the set of possible worlds where Kazimir Malevich was born in Ukraine to Polish parents. Sentence (2) refers to the set of possible worlds where Pablo Picasso was born in Spain to Spanish parents. These two sets are different because they are sets of possible worlds defined by different properties. Since the two sentences refer to two different sets of possible worlds, they also refer to different propositions, and have, therefore, different meanings.

The third problem was how to capture the logical relations between different declarative sentences. Propositions now allow us to capture these logical relations on the basis of corresponding quantitative relations between sets of possible worlds. Consider again sentences (3) and (4), repeated below.

(3) Kazimir paints

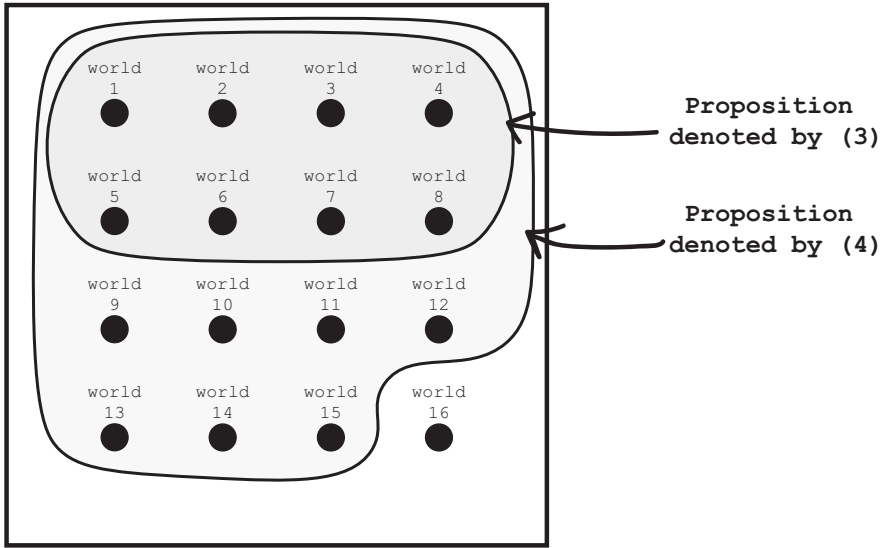
(4) Somebody paints

We have observed above that (3) logically entails (4), that is, whenever (3) is true, also (4) is. Consider now the possible worlds referred to by the two sentences in Table 11.1 from the previous section. Sentence (3), as we saw, refers to the set of possible worlds that includes worlds 1–8, that is, all possible worlds where Kazimir belongs to the set of painters. Sentence (4) refers to the set including possible worlds 1–15, that is, all possible worlds where at least one individual paints. The only world that is excluded from this set is world 16, which is the only world where nobody paints. By comparing the two sets, we observe that the set denoted by (3) is a subset of the set denoted by (4). This is exemplified in Fig. 11.5. As we see from the figure, the set of possible worlds that corresponds to the proposition expressed by (3) is a subset of the proposition expressed by (4). It is this quantitative relation between the two sets that legitimizes us to conclude that whenever (3) is true, also (4) must be. As every possible world that is included in the denotation of (3) is also automatically included in the denotation of (4), every state of affairs that makes (3) true makes (4) true as well.

We can easily generalize this pattern to all relations of entailment: For any two declarative sentences *A* and *B*, we say that *A* entails *B* if, and only if, the proposition referred to by *A* is a subset of the proposition referred to by *B*. This is the most significant outcome of the theory. It allows us to reduce the notion of logical entailment to a quantitative relation between two sets of objects, which is a type of relation we have already encountered when explaining other aspects of linguistic meaning. The only fundamental difference with what we have seen up to this point is that the objects we are now quantifying on are possible worlds.

This approach can easily be extended to other types of logical relations, for example, the relation of contrariety. We have seen above that (3) and (5), repeated below, are contrary sentences.

(5) Nobody paints



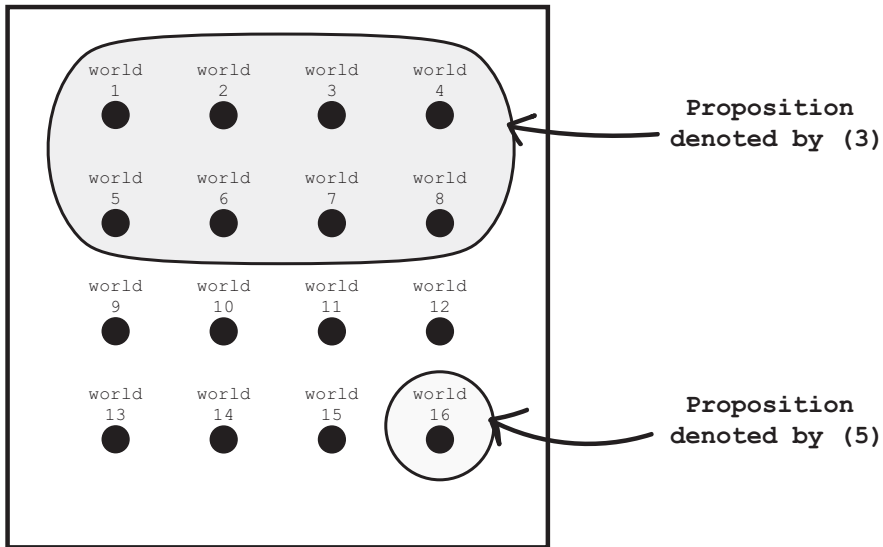
**Fig. 11.5** Each possible world in the model is represented by a dot. The larger set corresponds to the proposition expressed by sentence (4). The smaller set corresponds to the proposition expressed by (3)

In the terms of the current theory, we judge (3) and (5) as contrary because the sets of possible worlds referred to by the two sentences do not intersect with each other—that is, they do not have members in common. If we look, again, at Table 11.1, we find that sentence (3) refers to the set including possible worlds 1–8 whereas sentence (5) refers to the set including only possible world 16—the only possible world where no individual belongs to the set of those who paint. As we can appreciate from Fig. 11.6, the two sets do not overlap. This means that there is no logically conceivable possible state of the world that makes both sentences true at the same time. Hence, the two propositions are contrary propositions. Whenever one is true, the other must be false (though both of them could be false).

In this case as well, we can generalize the observation to all instances of logical contrariety: For any two declarative sentences *A* and *B*, *A* is the contrary of *B* if, and only if, the proposition referred to by *A* and the proposition referred to by *B* have an empty intersection.

## Possible Worlds and Compositionality

Possible worlds allow us to construct the notion of truth-conditions in our model and, with it, provide a significantly improved account of the meaning of declarative sentences. We must now consider how the theory of possible worlds can be adapted



**Fig. 11.6** As in Fig. 11.5, each possible world in the model is represented by a dot. The larger set corresponds to the proposition expressed by sentence (3). The smaller set corresponds to the proposition expressed by (5)

to the theory of compositionality we have developed in the previous chapters. That is, once we have replaced truth-values with propositions, we must also explain how propositions are derived compositionally from the meaning of the simple expressions that constitute them.

Consider again the sentence “Kazimir paints”. According to the account we have developed in the previous chapters, the meaning of the sentence is derived by combining the meaning of “Kazimir”, an individual object, with the meaning of “paints”, a set of objects. This set can be described by a corresponding characteristic function, which is symbolized in (9).

$$(9) \quad x \rightarrow 1 \text{ if } x \text{ paints, } 0 \text{ otherwise}$$

The function in (9) maps an object  $x$  into a truth-value, 1 or 0, depending on whether  $x$  paints. This function must now be changed, because, according to the framework we have developed in this chapter, we want declarative sentences to refer to propositions rather than truth-values. We can easily achieve the desired result by replacing (9) with (10).

$$(10) \quad x \rightarrow \text{the set of possible worlds where } x \text{ paints}$$

The function described in (10) is a function that maps an individual  $x$  into the set of possible worlds where  $x$  paints. When the individual Kazimir is applied to this function the function delivers, as its output, the set of possible worlds in the model where Kazimir paints, which is precisely the proposition we want for the sentence “Kazimir paints”.

We can apply the same strategy to all the denotations of the other intransitive verbs in the fragment of English that we are currently analyzing. We can also apply the same strategy to transitive verbs. For example, the verb “admires”, to which we assigned the function in (11), can now be assigned the function in (12).

(11)  $y \rightarrow (x \rightarrow 1 \text{ if } x \text{ admires } y, 0 \text{ otherwise})$

(12)  $y \rightarrow (x \rightarrow \text{the set of possible worlds where } x \text{ admires } y)$

Whereas (11) is a function from a first individual object  $y$  into a function from a second individual object  $x$  into a truth-value, (12) is a function that delivers, from the same inputs, a set of possible worlds, the set of those possible worlds where  $x$  admires  $y$ . The same goes for the denotations of determiners such as “every”, “some”, and “no”, which can also be adapted to the current analysis of propositions. Example (13) offers a revised version of the denotation of “every”.

(13)  $X \rightarrow (Y \rightarrow \text{the set of possible worlds where } X \text{ is a subset of } Y)$

We find that the use of possible worlds is fully compatible with a compositional semantics. In the next chapters, we will see that the analysis of propositions as sets of possible worlds brings with it a number of further advantages.

## References and Remarks

The notion of possible world should be traced back to Leibniz who, however, made only an informal use of it. It was with Carnap that possible worlds were adopted as part of a model-theory for the interpretation of a formal language (Carnap 1947). Carnap’s original framework underwent a number of improvements thanks to the work of a number of scholars. Kripke is without doubt amongst the most important (Kripke 1959a, b, 1962, 1963a, b, 1965) but not the only one. Other important contributors include Arnould Bayart (1958, 1959) Stig Kanger (Kanger 1957), Jaakko Hintikka (Hintikka 1961), Arthur N. Prior (Prior 1956, 1957), as well as Montague himself (Montague 1960).

Possible world semantics was originally developed as a framework for the interpretation of *modal* logic, that is, the logic that deals with the notions of contingency and necessity (we will return to these notions in the following chapter). In the decades that followed Carnap’s first formulation, possible world semantics found a number of applications in different domains. In the domain of natural language semantics, possible worlds were used to formalize, among other things, the distinction between *de re* and *de dicto* propositions (we will return also to this distinction in the next chapter), modal verbs (Kratzer 1981, 1991a), conditionals (Kratzer 1986, 1991b; Lewis 1973; Stalnaker 1968, 1991), and verbs of propositional attitude (we will return to this class of verbs in Chap. 12).

Despite its effectiveness, the notion of possible world has also been heavily criticized, particularly for its metaphysical implications. Amongst its strongest critics we should mention Willard Van Orman Quine (Quine 1956, 1960).



## References

- Bayart, A. (1958). Correction de la logique modal du premier et du second ordre S5. *Logique et Analyse*, 1, 28–45.
- Bayart, A. (1959). Quasi-Adéquation de la Logique Modal du Second Ordre S5 et Adéquation de la Logique Modal du Premier Ordre S5. *Logique et Analyse*, 2, 99–121.
- Carnap, R. (1947). *Meaning and necessity*. Chicago: University of Chicago Press.
- Hintikka, J. (1961). Modalities and quantification. *Theoria*, 27(3), 119–128.
- Kanger, S. (1957). Provability in logic. In *Acta universitatis stockholmiensis, Stockholm studies in philosophy* (Vol. 1). Stockholm: Almqvist and Wiksell.
- Kratzer, A. (1981). The notional category of modality. In H. J. Eikmeyer & H. Rieser (Eds.), *Words, worlds, and contexts* (pp. 38–74). Berlin/New York: de Gruyter.
- Kratzer, A. (1986). Conditionals. In A. M. Farley, P. Farley, & K. E. McCollough (Eds.), *Papers from the parasession on pragmatics and grammatical theory* (pp. 115–135). Chicago: Chicago Linguistics Society.
- Kratzer, A. (1991a). Modality. In A. V. Stechow & D. Wunderlich (Eds.), *Handbuch semantik/handbook semantics* (pp. 639–650). Berlin/New York: de Gruyter.
- Kratzer, A. (1991b). Conditionals. In A. V. Stechow & D. Wunderlich (Eds.), *Handbuch semantik/handbook semantics* (pp. 651–656). Berlin/New York: de Gruyter.
- Kripke, S. A. (1959a). A completeness theorem in modal logic. *Journal of Symbolic Logic*, 24(1), 1–14.
- Kripke, S. A. (1959b). Semantical analysis of modal logic (abstract from the twenty-fourth annual meeting of the association for symbolic logic). *Journal of Symbolic Logic*, 24(4), 323–324.
- Kripke, S. A. (1962). The undecidability of monadic modal quantification theory. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 8(2), 113–116.
- Kripke, S. A. (1963a). Semantical analysis of modal logic I. Normal modal propositional calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 9(5–6), 67–96.
- Kripke, S. A. (1963b). Semantical considerations on modal logic. *Acta Philosophica Fennica*, 16, 83–94.
- Kripke, S. A. (1965). Semantical analysis of modal logic II. Non-normal modal propositional calculi. In J. W. Addison, L. Henkin, & A. Tarski (Eds.), *Symposium on the theory of models* (pp. 206–220). Amsterdam: North-Holland.
- Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell Publishers.
- Montague, R. (1960). Logical necessity, physical necessity, ethics, and quantifiers. *Inquiry*, 3(1–4), 259–269.
- Prior, A. N. (1956). Modality and quantification in S5. *Journal of Symbolic Logic*, 21(1), 60–62.
- Prior, A. N. (1957). *Time and modality*. Oxford: Clarendon Press.
- Quine, W. V. O. (1956). Quantifiers and propositional attitudes. *The Journal of Philosophy*, 53, 177–187.
- Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.
- Stalnaker, R. (1968). A Theory of conditionals. *Studies in Logical Theory, American Philosophical Quarterly, Monograph*, 2, 98–112.
- Stalnaker, R. (1991). Indicative conditionals. In F. Jackson (Ed.), *Conditionals* (pp. 136–155). Oxford: Oxford University Press.

## Chapter 12

# Meaning and Necessity



### Necessary and Contingent

An important merit of the analysis of propositions as sets of possible worlds is that it captures a number of their core logical properties.

A first fundamental distinction in the domain of propositions is that between *necessary* and *contingent* propositions. A necessary proposition is a proposition that can never be false. Sentences (1) and (2) are examples of sentences expressing necessary propositions.

- (1) All painters are painters
- (2) Kazimir is Kazimir

The propositions expressed by these two sentences are not only true but also necessarily so. It is simply impossible to conceive of a logically consistent and coherent state of the world where all the individuals that are painters fail to satisfy the property of being painters. It is as impossible to conceive of a logically coherent and consistent state of the world where Kazimir is not identical to himself. The falsity of these sentences would defeat the fundamental laws of logic, as it would amount to claiming that there are objects that do satisfy a property  $P$  and, at the same time, do not satisfy the same property  $P$  and that an object is not identical to itself. Conversely, contingent propositions are propositions that have equal logical plausibility of being either true or false. Sentences (3) and (4) are examples of sentences expressing contingent propositions.

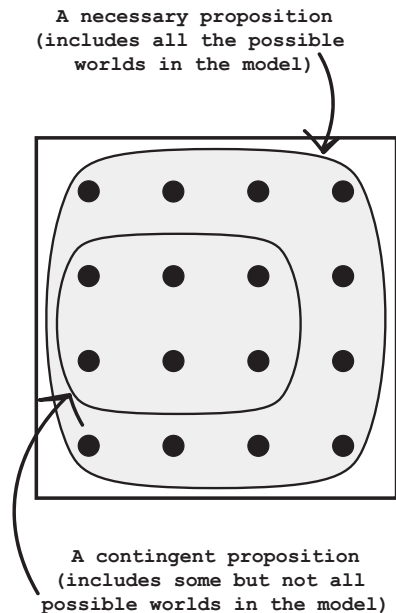
- (3) All painters are singers
- (4) Kazimir is a painter

It is as logically plausible to conceive of states of the world where all painters are singers as it is to conceive of states of the world where not all painters are singers. Similarly, it is as plausible to conceive of states of the world where Kazimir is a

painter as it is to conceive of states of the world where Kazimir is not a painter. The falsity of these sentences, hence, does not defeat the laws of logic.

The difference between necessary and contingent statements corresponds to a difference in truth-conditions, which we can readily formulate in a possible world semantics. As we just saw, necessary propositions are propositions that can never be false. Accordingly, in a possible world semantics, we can characterize necessary propositions as propositions that correspond to the set of *all* possible worlds in their model of interpretation. Sentence (1), to see an example, expresses a necessary proposition because it is true of all possible arrangements of properties and individuals in a model. It does not really matter who is a painter and who is not. Any possible assignment of this property to the individuals in the model is bound to satisfy the property that the set of painters is a subset of itself. Conversely, contingent propositions are propositions that can be either true or false. Accordingly, we can characterize contingent propositions as propositions that corresponds to a proper subset of the set of all possible worlds in a model. Sentence (3), to see an example, is contingent because not all logically admissible arrangements of properties and individuals in a model are bound to satisfy the condition that all painters are singers. There are always going to be possible worlds where all painters are singers as well as possible worlds where not all painters are singers. This difference is exemplified in Fig. 12.1.

**Fig. 12.1** The picture illustrates the difference in a possible world semantics between a necessary proposition, which corresponds to the set of all possible worlds in the model, and a contingent proposition, which corresponds to a subset of all the possible worlds in the model



## A Priori and A Posteriori

In the western philosophical tradition, the distinction between necessary and contingent propositions has long been taken to mirror a corresponding distinction in *epistemology*, the branch of philosophy that concerns itself with *knowledge*. It has been observed that sentient individuals acknowledge the truth of necessary and contingent propositions in two fundamentally different ways. The truth of necessary propositions is knowable a priori. The Latin term—which translates literally as “from what is before”—describes those propositions whose truth can be acknowledged without any experience of the facts they describe, but on the basis of pure reasoning. As an example, consider again sentence (1).

(1) All painters are painters

In order to tell whether the sentence is true we do not need to perform any particular observation or empirical research. We do not need to possess any knowledge of who is a painter and who is not. All we need to know to conclude that the sentence is true is that the relation expressed by the quantifier “all” is true whenever applied to two identical properties.

Contingent statements, conversely, are said to be a posteriori—literally, “from what is after”—because we can tell their truth only after we have verified empirically whether they succeed or fail in describing the actual facts. Consider again sentence (3).

(3) All painters are singers

We cannot tell whether (3) is true or false by simply considering the formal properties of the quantifier “all”. To do so, we must first acquire actual knowledge of whether it is the case that the set of painters is a subset of the set of singers and this is not something we can come to know on the basis of pure deductive reasoning.

The correlation between the concepts of necessary and a priori, on the one hand, and contingent and a posteriori, on the other, has been taken for granted in Western philosophy—more or less explicitly—until Immanuel Kant, who was among the first to provide serious arguments against it. The relationship between necessary and a priori, on the one hand, and contingent and a posteriori, on the other, will play an important role in part II. There, we will also encounter some remarkable arguments in favor of the Kantian insight that the two distinctions do not align with one another as neatly as it seems at a first glance.

## De Re and De Dicto

Another distinction we can formulate in a possible world semantics is that between de re and de dicto propositions. This distinction will also play an important role in part II. The Latin terminology is due to the medieval scholastic philosopher Thomas

Aquinas. *De re* literally means “about a thing”. Accordingly, *de re* propositions are propositions that are about a *res*—that is, an object. *De dicto* means “about what is said”. Hence, *de dicto* propositions are propositions that are about a description. To appreciate the distinction, consider the following pair of sentences.

- (5) Kazimir paints
- (6) The person wearing a hat paints

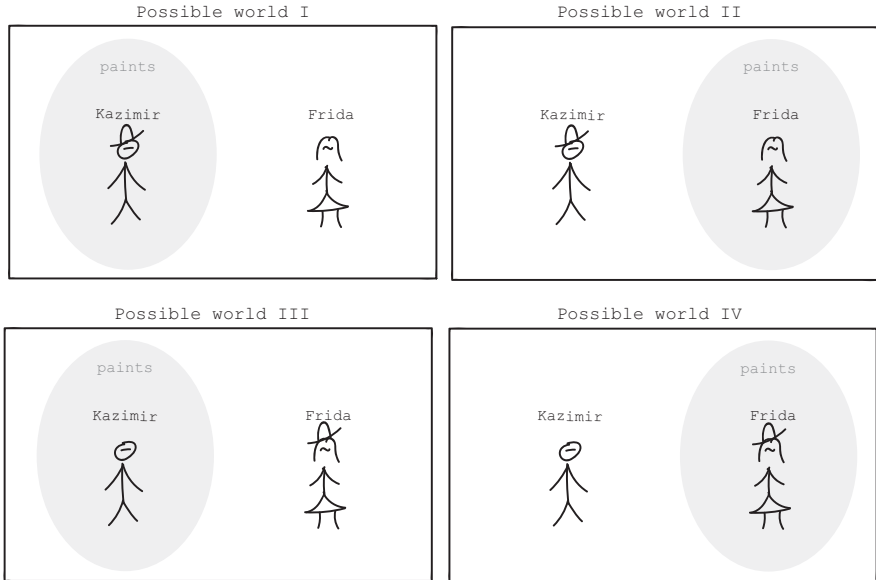
Sentence (5) expresses a *de re* proposition. It attributes a property—that of being painting—to a *res*—the object referred to by the proper name “Kazimir”. Sentence (6), conversely, expresses a *de dicto* proposition, which attributes the same property—that of painting—to a description—that expressed by the definite description “the person wearing a hat”.

In Chap. 10, we already considered the difference between proper names—such as “Kazimir”—and definite descriptions—such as “the person wearing a hat”. We saw that the two types of nominal expressions differ in that proper names denote their referent directly, whereas definite descriptions point to their referent through the mediation of a description. That is, whereas the meaning of a name is the object it refers to, the meaning of a definite description is a property that has the potential to uniquely identify an object. Whereas names tell us *what* we are talking about, definite descriptions tell us *how* we are talking about it.

The difference between names and descriptions is, of course, at the basis of the distinction between *de re* and *de dicto* propositions. Thanks to possible worlds, we can now characterize precisely how proper names and descriptions contribute to these two different types of propositions. Let us consider a very simple model constructed around two individuals and two properties. The two individuals are Kazimir and Frida and the two properties are that of painting and that of wearing a hat. Consider then the four possible worlds in Fig. 12.2, corresponding to various combinations of these objects and properties. Possible worlds I and III are worlds where Kazimir paints but Frida does not. Possible worlds II and IV are worlds where Frida paints but Kazimir does not. Furthermore, whereas Kazimir is wearing the hat in worlds I and II, Frida is wearing the hat in worlds III and IV.

Consider now which possible worlds correspond to the propositions expressed by (5) and (6). Sentence (5), as we know, refers to the set of possible worlds where the individual Kazimir belongs to the set of those who paints. In the model we are considering, this set includes possible worlds I and III, as those are the possible worlds where Kazimir paints, and excludes possible worlds II and IV, as those are the possible worlds where Kazimir does not paint. Sentence (6), instead, refers to the set of possible worlds where the unique individual who wears a hat also belongs to the set of those who paint. In the context of our model, this set includes possible worlds I and IV and excludes possible worlds II and III. This is because only in I and IV the unique individual who uniquely satisfies the property of wearing a hat also satisfies the property of painting.

We see, from this simple example, that the two sentences refer to two partially overlapping, yet different propositions. Sentence (5) refers to the set of possible worlds where Kazimir paints. Sentence (6) refers to the set of possible worlds where



**Fig. 12.2** A model of four possible worlds comprising different possible arrangements of two individuals—Kazimir and Frida—and two properties—the property of painting and that of wearing a hat

whoever happens to be the person wearing the hat paints. Since in some possible worlds Kazimir is not the person wearing the hat, there are possible worlds that belong to one proposition but not to the other.

We understand now more precisely how proper names and definite descriptions contribute to different types of propositions. Whereas “Kazimir” always refers to Kazimir, “the person with a hat” may refer to Kazimir or Frida in different possible worlds, depending on who wears the hat. Hence, whereas the reference of a proper name is enforced as a matter of necessity, the reference of a description is individuated contingently. Using a term introduced by Kripke, we will say that proper names are *rigid designators*, meaning that they are expressions whose reference remains constant across different possible worlds. Definite descriptions, in contrast, are not rigid designators. Their reference is decided indirectly through the mediation of a description, hence, they may pick different referents in different possible worlds, depending on which individual happens to satisfy the relevant description.

## References and Remarks

Modality—and with it the notions of necessity and contingency—entered the forum of philosophy with Aristotle (see, in particular, the *Prior Analytics*). Modality was then a central topic of investigation during the Middle Ages amongst Western as

well as Arabic philosophers. It was in this period that scholars developed the notion that modality involves reference to non-actual alternatives to the actual world. For a thorough excursus into the development of modal theorizing in the middle ages we refer the reader to Simo Knuuttila's *Modalities in Medieval Philosophy* (Knuuttila 1993). As we saw at the end of the previous chapter, it was only much later that the notion of a possible world was entertained and a formal semantics for the interpretation of modal statements, of the type we have used in this chapter, was formulated.

The distinction between a priori and a posteriori as two opposing modes of justification of knowledge is made and discussed by Kant in his *Kritik der reinen Vernunft*. As we already mentioned in this chapter, there are reasons to doubt the overlap between the notions of necessary and a priori, on the one hand, and contingent and a posteriori, on the other. Some of these reasons are already discussed by Kant. We will return to this issue in several occasions in Part II. In Chap. 19, we will see how the notions of necessary and a priori have been exploited to provide an epistemology of mathematical propositions. Then, in Chaps. 21 and 23, we will review the case of statements that express necessary propositions but are not a priori and the case of statements that express contingent propositions but are nonetheless a priori.

A further distinction, which we did not review in this chapter, is that between *synthetic* and *analytic* statements. Analytic statements are statements whose truth depends solely on the meaning of the terms in which they are formulated. Consider, as an example, the sentence "all cars are vehicles". That the sentence is true follows mechanically from the meaning of the words that occur in it. As it is part of the definition of the word "car" that anything that is a car is also a vehicle, the truth of the sentence follows mechanically from such definition. Synthetic statements, conversely, are statements whose truth depends on the facts of the world they aim to describe. Consider, as an example, the sentence "all cars are fast". Whether the sentence is true depends on whether it is really the case that all cars are fast. Again, it is intuitive to think of the analytic/synthetic as overlapping with both the a priori/a posteriori distinction and the necessary/contingent distinction. Against the conflation of these notions, Kant famously discussed the case of mathematical statements, which, according to him, are both necessary and a priori but nonetheless synthetic. In the twentieth century, the notion of analyticity and its relation to the a priori and the necessary has been heavily criticized by Quine, starting from his 1951 article "Two Dogmas of Empiricism" (Quine 1951).

The terms *de re* and *de dicto* are due to Thomas Aquinas, who adopted them for the first time in his *De Propositionibus Modalibus* although the distinction had already been observed by Aristotle. The terms were introduced to the forum of contemporary modal logic by Georg Henrik von Wright who adopted them in his 1951 essay in modal logic (von Wright 1951).

The notion of rigid designation is introduced by Kripke in his *Naming and Necessity* (Kripke 1980) with the help of a number of examples. A particularly famous one concerns the proper name "Nixon" and the description "the U.S. President in 1970". As Kripke notices, whereas it is perfectly conceivable that Nixon may not have been the U.S. President in 1970, it is inconceivable that Nixon

may not have been Nixon. The example shows how the description “the U.S. President in 1970” may change its reference across different possible worlds whereas the proper name is a rigid designator whose reference remains constant across different possible worlds. We will return to some of the arguments made by Kripke in the same text when we will discuss the description theory of proper names in Chap. 22.

## References

- Knuuttila, S. (1993). *Modalities in medieval philosophy*. London/New York: Routledge.
- Kripke, S. A. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.
- Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review*, 60, 20–43.
- von Wright, G. H. (1951). *An essay in modal logic*. Amsterdam: North-Holland Publishing Company.



# Chapter 13

## Meaning and Attitudes



### Sentences Within Sentences

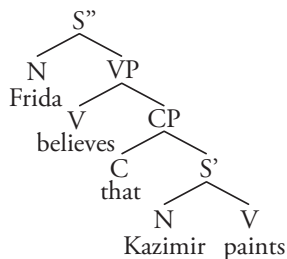
As we saw at the end of Chap. 11, a positive feature of the view of propositions as sets of possible worlds is that it is fully compatible with the principle of compositionality. Propositions, as expressed by declarative sentences, can be derived by combining the meanings of the simple parts constituting the sentence through the mechanism of functional application. In this chapter, we will consider a further advantage of this view of propositions for the principle of compositionality. Propositions can themselves contribute to more complex meanings. This is what happens, for example, with a sentence such as (1).

(1) Frida believes that Kazimir paints

The expression in (1) is a declarative sentence, although of a peculiar sort. Differently from the declarative sentences we have considered so far, it contains a declarative sentence amongst its constituents. This is the declarative sentence “Kazimir paints”.

In a phrase-structure grammar, sentence (1) is assigned the structure in Fig. 13.1. According to this structure, the sentence is constructed by first combining “Kazimir” and “paints” to obtain the declarative sentence ( $S'$ ) “Kazimir paints”. This sentence is then combined with the complementizer “that” to obtain the complementizer phrase (CP) “that Kazimir paints”, which is then combined with the verb “believes” to obtain the verb phrase (VP) “believes that Kazimir paints”. Finally, the verb phrase is combined with the proper name “Frida” to obtain the declarative sentence ( $S''$ ). The most notable feature of this structure is that the node of syntactic type  $S$  appears twice in it—as  $S'$  and  $S''$ . This structure is *recursive*, as it represents a syntactic object that contains a syntactic object of its very same type among its constituents. In what follows, we will refer to the most embedded  $S'$  node in the structure of (1) as the *complement* sentence—because it acts as the

**Fig. 13.1** Phrase-structure of the sentence “Frida believes that Kazimir paints”



direct complement of the verb “believe”. We will refer to the higher  $S''$  node as the *matrix* sentence. We have already referred to the expression “that” as a *complementizer*. This term expresses the fact that the grammatical function of a complementizer is to turn an independent sentence into the complement of a verb. For simplicity’s sake, in what follows, we will disregard the complementizer from the analysis of this type of structures. The reason for doing so is that the complementizer can be regarded—at least at the level of analysis relevant to our discussion in this chapter—as performing a purely syntactic function but not a semantic one. That is, the complementizer plays a grammatical role but does not contribute a meaning of its own. Finally, we will refer to verbs such as “believe” as verbs of *propositional attitude*. The reason behind this term is the topic of the next section.

## Propositions Within Propositions

The term *propositional attitude*, which we used to refer to verbs such as “believe”, reflects the idea that the recursive syntax of sentences such as (1) corresponds to a recursive semantics. As, syntactically, (1) corresponds to a sentence within a sentence so, semantically, it expresses a proposition about a proposition. Intuitively, (1) expresses the proposition that Frida believes the proposition that Kazimir paints.

We can regard verbs of propositional attitudes as expressing relations, although relations of a special sort. We have discussed the concept of relation in Chaps. 8 and 9, when we discussed transitive verbs—such as “admire” and “despise”—and determiners—such “every” and “no”. Both these classes of expressions denote relations—relations between individuals in the case of transitive verbs; relations between sets in the case of determiners. We will see now that verbs of propositional attitude can also be analysed as expressing relations although relations between individuals and propositions. Relations of this sort are called of propositional attitude because they reflect the cognitive attitude that an individual has towards the content of a proposition.

This view was formalized in a model-theory for the first time by the philosopher and logician Jaakko Hintikka. In Hintikka’s framework, a verb such as “believe” is taken to express the following function.

(2)  $p \rightarrow (x \rightarrow \text{the set of possible worlds where } BEL_x \subseteq p)$

In a nutshell, (2) expresses a relation between a proposition  $p$  and an individual  $x$  which is true just in case what  $x$  believe to be true entails  $p$ . In the formalization provided by Hintikka, this relation is modelled as a function that takes an input value  $p$  to deliver another function that takes an input value  $x$  to deliver a set of possible worlds. The lower-case letter  $p$  is a variable over *propositions*, which are, in our framework, sets of possible worlds. The variable tells us that the first input of the function is a proposition. The lower-case  $x$ , as we know from Chap. 6, is a variable over individual objects. It tells us that the second input of the function is an individual object. The output of the function is a set of possible worlds where the values we have chosen for  $x$  and  $p$  satisfy the following relation:

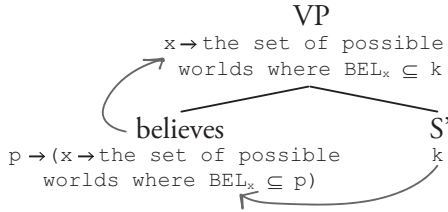
(3)  $BEL_x \subseteq p$

To understand the condition imposed by (3) on  $x$  and  $p$ , we must first understand the expression  $BEL_x$ . Simply put,  $BEL_x$  is a short hand for “what  $x$  believes”. More formally,  $BEL_x$  is the proposition that correctly describes what  $x$  believes to be the case. This proposition is equivalent to a set of possible worlds—the set of possible worlds that are consistent with what the individual  $x$  believes to be the case. The condition in (3), then, requires that  $BEL_x$  is a subset of  $p$ —that is, that the set of possible worlds that are consistent with what  $x$  believes is a subset of the possible worlds where  $p$  is true.

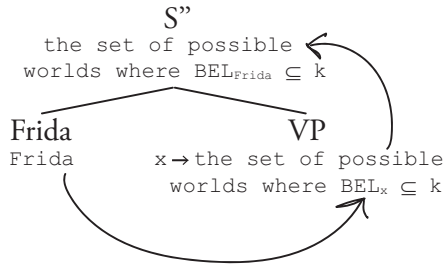
To see Hintikka’s analysis of “believe” at work, let us apply it to sentence (1), repeated below.

(1) Frida believes that Kazimir paints

As we saw above, the sentence is constructed by first combining “Kazimir” and “paints” in the declarative sentence “Kazimir paints”, by then combining this sentence with the verb “believes” (through the mediation of the complementizer “that”, which we are, however, disregarding for the reasons mentioned above), and, finally, by combining the resulting verb phrase with the proper name “Frida”. We already know from the previous chapters that the sentence “Kazimir paints” refers to a proposition—the set of possible worlds where Kazimir belongs to the set of those who paint. For convenience sake, let us symbolize this proposition with the letter  $k$ . To obtain the meaning of (1), we must first derive the meaning of the verb phrase “believes that Kazimir paints” as the result of combining the meaning of “believes” with the meaning of “Kazimir paints”. As established in (2), “believes” refers to a function that takes a proposition as its input argument. “Kazimir paints” denotes a proposition. Hence, the meaning of the verb phrase “believes that Kazimir paints” can be derived as the result of applying  $k$ —the proposition referred to by “Kazimir paints”—to the function referred to by “believes”. This first step is illustrated in Fig. 13.2. The meaning that results for the verb phrase “believes that Kazimir paints” is yet another function—a function that takes an individual  $x$  as its input and delivers, as its output, the set of possible worlds where  $BEL_x$ —what  $x$  believes—is a subset of  $k$ . The final step to obtain the meaning of (1) requires, therefore, comb-



**Fig. 13.2** The figure illustrates the compositional derivation of the meaning of the Verb Phrase (VP) “believes that Kazimir paints”. “Believes” denotes the function  $p \rightarrow (x \rightarrow \text{the set of possible worlds where } BEL_x \subseteq p)$  whereas the sentence  $S'$  “Kazimir paints” denotes a proposition, symbolized in the figure as  $k$ . Application of  $k$  to the function denoted by “believes” delivers the function  $x \rightarrow \text{the set of possible worlds where } BEL_x \subseteq k$  as the meaning of the VP

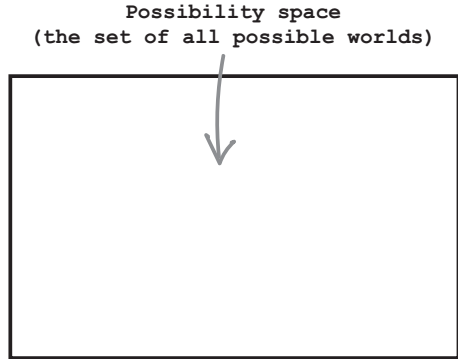


**Fig. 13.3** The figure illustrates the compositional derivation of the meaning of the sentence ( $S''$ ) “Frida believes that Kazimir paints”. “Frida” refers to the individual Frida whereas the VP “believes that Kazimir paints” denotes the function  $x \rightarrow \text{the set of possible worlds where } BEL_x \subseteq k$ . Application of Frida to this function delivers the set of possible worlds where the condition  $BEL_{Frida} \subseteq k$  is met as the reference of  $S''$

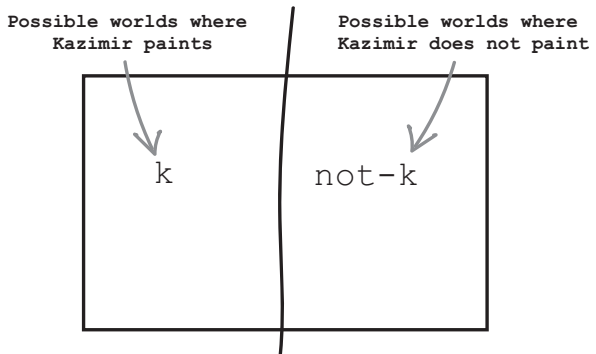
ing the meaning of the verb phrase “believes that Kazimir paints” with the meaning of the proper name “Frida”. This second step is illustrated in Fig. 13.3. The result is a proposition—the set of possible worlds where what Frida believes entails that Kazimir paints. According to these truth-conditions, the sentence is true if, and only if, the set of possible worlds that represent what Frida believes is a subset of the set of possible worlds where Kazimir paints.

More intuitively, we can think of (1) as informing us about the location of Frida’s beliefs in the space of logical possibilities. Let us visualize the space of all logical possibilities as a rectangular space, as in Fig. 13.4. Let us then divide this space into two subspaces corresponding, respectively, to the possibilities where Kazimir paints and those where Kazimir does not paint, as illustrated in Fig. 13.5. What sentence (1) does is to inform us that Frida’s belief is located within the portion of the possibility space that is consistent with proposition  $k$ , as illustrated in Fig. 13.6. From the truth of (1), we know that, whatever the proposition that describes the world as Frida believes it to be, such a proposition is logically consistent with the proposition that Kazimir paints.

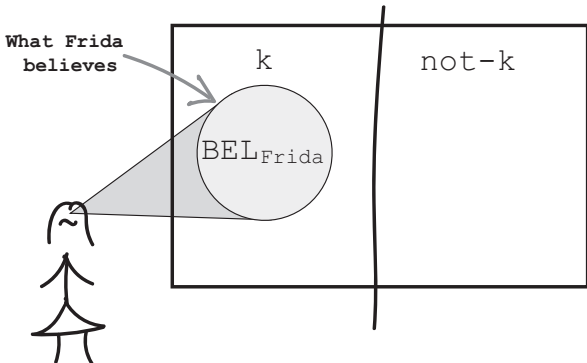
**Fig. 13.4** A visualization of the possibility space—the set of possible worlds



**Fig. 13.5** The possibility space is divided in two areas: to the left, all the possible worlds where Kazimir paints and, therefore, verify proposition  $k$ ; to the right, all the possible worlds where Kazimir does not paint and, therefore, falsify proposition  $k$



**Fig. 13.6** The sentence “Frida believes that Kazimir paints” informs us that the set of possible worlds that are consistent with what Frida believes (symbolized as  $BEL_{Frida}$ ) is included within the region of the logical space that verifies proposition  $k$



## Modal Bases

Hintikka’s analysis formalizes the idea that “believe” expresses a relation between an individual and a proposition and does so in a fully compositional fashion. Hintikka’s semantics for “believe” can also be easily generalized to the other verbs of propositional attitude—such as “think”, “know”, “remember”, “say”, “suspect”,

“hope”, “wish”, “deny”, “imagine”. The idea is simple. As we saw above, “believe” relates what a subject  $x$  believes to a proposition  $p$ . We symbolized what  $x$  believes as  $BEL_x$ . We will now refer to this proposition as the *modal base* of the verb “believe”—the proposition that defines the portion of the possibility space that is consistent with what  $x$  believes to be the case. Different verbs of propositional attitude share the same common semantics, but they differ in their modal bases. Consider, as an illustrative example, the verb “hope”:

(4) Frida hopes that Kazimir will paint

The verb “hope” also expresses a relation between a modal base and a proposition. However, this time the modal base is not the set of possible worlds that are consistent with what the subject believes. Rather, it is the set of possible worlds that are consistent with what the subject *hopes*. Therefore, sentence (4) is true if, and only if, the proposition that is consistent with what Frida hopes entails that Kazimir paints. The same rationale can be extended to the other verbs of propositional attitudes, by identifying the modal base that is relevant to each verb.

## Attitudes De Re and De Dicto

Hintikka’s analysis expresses propositional attitudes as relations of entailment between a subject’s modal base and a proposition. This relation of entailment is, in turn, represented in the model as the subset relation between two sets of possible worlds. A considerable advantage of this analysis is that it captures a number of logical properties of propositional attitudes. One that will become especially relevant in part II is the distinction between de re and de dicto propositional attitudes.

In the previous chapter, we distinguished de re propositions—such as the one referred to by (5)—from de dicto propositions—such as the one referred to by (6).

(5) Kazimir paints

(6) The person wearing a hat paints

De re propositions are propositions about a *res* whereas de dicto propositions are about a description. The proposition expressed by (5) is a de re proposition because it corresponds to the set of possible worlds where Kazimir paints. The proposition expressed by (6) is a de dicto proposition because it corresponds to the set of possible worlds where whoever happens to be wearing the hat paints.

We can now capture the same distinction in the context of propositional attitudes. Sentence (1), repeated below, expresses a de re propositional attitude because the object of Frida’s belief is a de re proposition.

(1) Frida believes that Kazimir paints

The sentence reports a belief that Frida has about Kazimir. It is true if, and only if, what Frida believes entails the de re proposition that Kazimir, rigidly identified, paints. Sentence (7), conversely, expresses a de dicto propositional attitude.

## (7) Frida believes that the person wearing a hat paints

The sentence reports a belief that Frida has towards a *de dicto* proposition. It is true if, and only if, what Frida believes is consistent with the set of possible worlds where whoever happens to be wearing the hat in those possible world paints.

## References and Remarks

Verbs of propositional attitudes are important to linguistic theory because they showcase a core property of natural language syntax and semantics: *recursion*. Natural language allows its users to construct and interpret complex expressions that contain within themselves expressions of the same syntactic and semantic type. Verbs of propositional attitudes, for example, allow us to produce sentences that contain sentences within themselves, which, in turn, are interpreted as propositions that are themselves about propositions. Recursion has been identified, since the early days of modern generative grammar, as the source of the potentially infinite generative capacity of natural language grammars. It is because of the recursive nature of their combinatorial rules that they allow us to produce expressions of potentially infinite length and complexity. This can be readily exemplified with verbs of propositional attitudes. Consider again the sentence “Frida believes that Kazimir paints”. As we saw, it is a sentence that contains a sentence within itself. This sentence can in turn be embedded within another sentence by means of a verb of propositional attitudes, such as “Pablo thinks that Frida believes that Kazimir paints”. The resulting sentence can in turn be embedded within an even larger sentence, such as “Alexandra said that Pablo thinks that Frida believes that Kazimir paints”. In principle, this procedure could continue infinitely. For a recent discussion on the notion of recursion in the context of natural language we recommend Steven Pinker and Ray Jackendoff’s article “The faculty of language: what’s special about it?” (Pinker and Jackendoff 2005).

On the same note, we should also observe that things are more complicated on the semantic side of recursion. Sentences that are embedded within verbs of propositional attitudes have different semantic properties than those that are not. Consider first the following inference:

Superman can fly  
 Superman is identical to Clark Kent

---

Clark Kent can fly

Speakers of English judge the inference as a valid one—that is, they agree that the truth of the conclusion “Clark Kent can fly” follows deductively from the truth of the two premises “Superman can fly” and “Superman is identical to Clark Kent”. What enables the inference is the principle of indiscernibility of identicals, also known as Leibniz's Law. According to this principle, if any two objects *a* and *b* are

identical, then any property that is true of  $a$  is also true of  $b$  and vice versa. Simply put, identical objects have identical properties. The principle enables us to replace the referring expression “Superman” in the sentence “Superman can fly” with the coreferential expression “Clark Kent” without affecting the truth of the sentence. Consider, however, the following inference:

Kazimir believes that Superman can fly  
 Superman is identical to Clark Kent

---

Kazimir believes that Clark Kent can fly

The inference is judged as non-valid. The truth of the premises does not grant the truth of the conclusion. The main difference, compared to the previous inference, is that the sentences “Superman can fly” and “Clark Kent can fly” are now embedded within the scope of a verb of propositional attitude. In such context Leibniz’s Law does not apply—we cannot replace a referring expression with a different one with the same reference without affecting the truth of the sentence. A context where expressions with identical referents cannot replace one another *salva veritate* is called an *opaque* context. The issue of opaque contexts was first observed by Frege in his “Über Sinn und Bedeutung” (Frege 1892) but was formulated as a problem in formal logic by Quine in his “Quantifiers and Propositional Attitudes” (Quine 1956). We will return to a number of issues concerning identity and opacity in part II of the book.

The semantics of verbs of propositional attitude we have presented in this chapter is based on that originally devised by Jaakko Hintikka (Hintikka 1962, 1969). Compared to Hintikka’s original formulation, ours is simplified in an important respect. Whereas we defined a modal base as a function from individuals into a proposition, in Hintikka’s formulation a modal base is a function from individuals and possible worlds into propositions. The modal base for the verb “believe”, to see an example, is more appropriately symbolized as  $BEL_{x,w}$ —the function that takes an individual  $x$  and possible world  $w$  to return the proposition that is consistent with what  $x$  believes to be the case in  $w$ .

Hintikka’s semantics analyzes verbs of propositional attitude as universal quantifiers over possible worlds. This analysis is attractive for its simplicity but has also been criticized for a number of reasons. One that is particularly relevant to us is that it fails to fully capture the meaning of *de re* propositional attitudes. We will return to this problem and to its possible solutions in Chaps. 25 and 26.

A note on complementizers is in place. In this chapter, we have maintained that complementizers perform an exclusively syntactic function and do not contribute to the meaning of the complex expressions in which they occur. In favor of this view, we can observe that complementizers can often be omitted. However, it should also be observed that complementizer omission—or “deletion”, as it is canonically referred to in the syntactic literature—is possible only with certain embedding verbs and is subject to significant cross-linguistic variation. Furthermore, a point has been made that a comprehensive analysis of the condition that license the omission of complementizers cannot just focus on the syntactic properties of the structures in



which they occur but must also consider its semantic properties (Bianchi and Frascarelli 2017). This suggests that complementizers may have a semantic value after all. We refer the reader interested in this issue to the comprehensive overview by Valentina Bianchi and Mara Frascarelli (Bianchi and Frascarelli 2017). It is also worth pointing out that there are complementizers with an explicit semantic component. Relevant examples in English are “if”, which expresses a logical relation between the main clause and the embedded clause, and “when”, which expresses a temporal relation.

## References

- Bianchi, V., & Frascarelli, M. (2017). Complementizer deletion. In M. Everaert & H. C. van Riemsdijk (Eds.), *The Wiley Blackwell companion to syntax* (Vol. 2, 2nd ed., pp. 963–992). Oxford: Wiley-Blackwell.
- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100(1), 25–50.
- Hintikka, J. (1962). *Knowledge and belief: An introduction to the logic of the two notions*. Ithaca, NY: Cornell University Press.
- Hintikka, J. (1969). Semantics for propositional attitudes. In J. Davis et al. (Eds.), *Philosophical logic*. Dordrecht: D. Reidel.
- Pinker, S., & Jackendoff, R. (2005). The faculty of language: What’s special about it? *Cognition*, 95, 201–236.
- Quine, W. V. O. (1956). Quantifiers and propositional attitudes. *The Journal of Philosophy*, 53, 177–187.

# Chapter 14

## Natural Language Metaphysics



### Quantity and Quality

With this chapter, we conclude the first part of our book and, with it, our introduction to the main tenets of Montague's model-theory of natural language meaning. Indeed, we have only touched upon the features of the theory that are most essential to it and we have only discussed a handful of the grammatical categories that the theory can account for: proper and common names, determiners, transitive and intransitive verbs, verbs of propositional attitude, and the declarative sentences that can be produced by combining these elements. We hope, nonetheless, that this brief overview has provided a concrete understanding of the principles and mechanisms at the basis of the theory. These principles and mechanisms will provide us with a general framework for the discussion to come. In fact, with Montague's lesson in mind, we shall now return to our initial question—what is meaning?

There are two main insights we can draw from Montague's experiment. The first concerns the *quantitative* properties of meaning. The second concerns its *qualitative* ones. As we saw, Montague's theory is driven by the idea that, by looking at meaning from a purely *formal* perspective, we can meet Frege's principle of compositionality and, in this way, reconcile the fact that words have individual meanings with the fact that these meanings interact combinatorially to produce more complex meanings, in harmony with the workings of the phrase-structure grammar of natural language. To achieve this goal, the theory characterizes the notion of meaning in terms of a *homomorphism* between two distinct domains of objects: on the one hand, the simple expressions of the language, organized around their grammatical properties and relations; on the other hand, the referents of these simple expressions, organized as a model of objects. The model provides the world (or universe) of interpretation of the language. The particular model adopted by Montague is constructed around two primitive formal notions: the notion of an object and the notion of a set. Both notions are intended in an abstract sense. All that

matters for something to qualify as an object in the model is that it can be conceived as a *mathematical quantity*—that is, something that can be assigned membership to a set. Sets are, in fact, the simplest mathematical structures that relate these quantities. These two notions alone allow us to construct mathematical structures of higher and higher complexity, such as functions, functions of functions, functions of functions of functions, and so on. In turn, these mathematical structures allow us to mirror the compositional properties of phrase-structure grammar in the domain of meaning. As Malevich decomposed the image to its formal ingredients, so Montague’s theory uncovers the quantitative edges of meaning—the algebraic properties that meaning must meet in order to become the source and product of the creative machinery of language.

The second lesson we draw from Montague is of a qualitative nature. In the last three chapters, we argued that the model of interpretation of natural language also entails a dimension of possibility. The different possible arrangements that the objects in a model may assume are themselves objects of quantification. We called these peculiar objects possible worlds. Possible worlds allow us to define the notion of proposition and, with it, to assign a meaning to declarative sentences that explains their logical properties as well as the way they contribute to the meaning of larger expressions. By introducing possible worlds, however, we also ended up introducing a *qualitative* distinction across the objects in the model. We must now distinguish between two different classes of objects in the model: individual entities—regular objects such as Kazimir, Frida, and the die—and possible worlds—hypothetical arrangements of the regular objects. To appreciate this point more concretely, consider again the denotation that we provided for the intransitive verb “paints” at the end of Chap. 11:

(1)  $x \rightarrow$  the set of possible worlds where  $x$  paints

As we saw, this is a function that takes an individual entity  $x$  as its input value and delivers a set of possible worlds as its output value—the set of possible worlds where  $x$  paints. In describing this function, we must specify the type of objects that qualify as its input as well as the type of objects that qualify as its output. That is, it is not enough to say that (1) is a function from objects into sets of objects. We must also specify their *type* and, more precisely, whether these objects are individual entities or possible worlds. If this is correct, we must conclude that the model of interpretation of natural language meaning is not only sensitive to the purely quantitative properties of its objects, but also to some qualitative distinctions.

## Natural Language Metaphysics

Consider Kazimir and Frida, two objects in our model. They are qualitatively different objects characterized by distinct properties. Yet, in the context of the specific mechanics of our theory, this difference is irrelevant. As far as the theory is concerned, both Kazimir and Frida belong to the same *class* of entities as they can

perform the same function in the context of the theory. For example, they can both act as input values to the function in (1). A possible world and Frida, on the other hand, are not just different objects on general grounds. They are different objects also from the perspective of the very logic of the model of interpretation, as they perform different compositional functions. A possible world, for example, cannot act as the input value to the function in (1) but a set of possible worlds can be the output of that same function.

We will refer to the practice of identifying and studying the qualitative differences between the objects in the model that are relevant to the interpretation of natural languages as *natural language metaphysics*, a label originally proposed by the linguist Emmon Bach. Why metaphysics? Metaphysics, as is generally understood, is the branch of philosophy that investigates the nature of reality in the most general possible sense. Metaphysicians are interested in discovering the properties of the world and its constituents that are universal and immutable or, adopting the terminology of philosophers, *essential*. Essential properties are to be understood in opposition to *accidental* ones. Consider again the two individuals Kazimir and Frida. No doubt, as we already observed, they are distinct individuals. They have different heights, weights, ages, nationalities, genders, personalities, friends, interests, tastes, attitudes, and so on. Furthermore, they are also two physically distinct objects, made of different physical particles, occupying different spatial coordinates. In fact, one may argue that, from a purely material perspective, they are completely different objects down to the smallest physical particles. Yet, despite their differences, it is natural to regard them as sharing common properties. For example, they are both human beings. But how can we regard them as belonging to a common category when we have just concluded that they are distinct down to the smallest physical units that constitute them? The answer given by those philosophers who, like Aristotle, adopt the notion of essential property is that being human is an essential property of Frida and Kazimir. It is a property that belongs to them in a way that holds independently of the contingent and material circumstances that affects them. The distinction between accidental and essential properties is often related to the distinction between *contingent* and *necessary* properties. We have introduced the distinction between contingent and necessary propositions in Chap. 12. The same distinction applies to properties. Contingent properties are properties that an object happens to have but could also not have had. Necessary properties are properties that an object must have as a matter of necessity and, therefore, could not not have had. The purpose of metaphysics is, according to this view, that of identifying the properties of reality that are necessary and, therefore, essential to it. The purpose of natural language metaphysics is, therefore, similar to that of philosophical metaphysics, although its scope is restricted to a different domain of objects. Whereas proper metaphysics is concerned with the essential properties of reality as a whole, natural language metaphysics is concerned with the essential properties that characterize the model of interpretation of natural language.

Interestingly, the theory we have developed in the previous chapters can already tell us something about the metaphysics of natural language meaning, what is essential about it. It suggests that the model of natural language interpretation is

characterized by two essential properties: the property of being an individual entity and the property of being a possible world. These properties distinguish in turn two essential classes of objects in the model: individual entities and possible worlds. These two types of objects share common quantitative properties but have different qualitative profiles. Their difference is not one that has to do with the mechanical functioning of the compositional machinery that generates meanings. It has not to do with the logical relations that enable the mechanism of functional application. It has not to do with the syntax of the system. It rather has to do with the distinction between things as they are and things as they can be, which is a distinction of an altogether different sort and worth reflecting upon.

## Quantity Before Quality

As a rule of the thumb, semanticists are always very considerate in introducing new essential properties in their model-theoretic accounts of natural language meaning. The reason has to do with a widely adopted principle of scientific conduct known as *Occam's razor*. Occam's razor is the principle—attributed to the medieval philosopher William of Ockham—whereby, whenever there is more than one explanation available for a certain phenomenon, the *simplest* explanation should always be preferred. A quantitative approach to meaning is simpler than a qualitative one. It explains meaning on the basis of the ingredients already available in the model, without introducing new qualitative distinctions. Hence, a quantitative approach should always be preferred.

This said, there are cases in which introducing a new essential property in the model seems the only practicable way to capture the meaning of certain expressions. This is what we saw in the previous chapters, where we considered some motivations for the introduction of possible worlds as the best way to characterize the notion of proposition and, in this way, express the meaning of declarative sentences and verbs of propositional attitudes. In the course of time, semanticists have introduced more and more essential categories demonstrating a rather rich array of qualitative distinctions within the fabric of meaning. It is worth briefly considering at least two such essential categories, *points in time* and *events*, as they have both played a significant role in the semantic theorizing of the last decades.

## Time

In many languages, including English, verbs are marked for *tense*. Compare the following two sentences:

- (2) Kazimir paints
- (3) Kazimir painted

The suffix “-s”, attached to “paint” in sentence (2), indicates that the verb is in the present tense. The suffix “-ed”, attached to the same verb in (3), indicates past tense. This difference in grammatical form between the two sentences corresponds to a difference in meaning. Both sentences express the fact that Kazimir has the property of painting. However, whereas according to (2) Kazimir satisfies the property in the present, according to (3) he did so at some point in the past.

The most convenient way to express the semantics of tense in a model-theory requires enriching the model with yet another dimension and a corresponding new class of objects: *points in time*. By introducing points in time as objects of quantification in the model of interpretation of natural language, we can provide a straightforward account of the meaning of the different tenses. As a matter of fact, points in time were already included as an independent essential category—together with possible worlds and individual entities—in Montague’s original formulation of his theory. It is difficult to see how tense could be explained otherwise, without enriching the model with a temporal dimension, and relying exclusively on individual entities and possible worlds.

## Events

Another essential category that has played an important role in semantic theorizing is that of *events*. There is a lively ongoing philosophical discussion concerning how to exactly characterize the class of objects we call events. For the sake of our discussion, we can think of the difference between the sort of individual entities we have considered so far and events as a difference between things that *are* and things that *happen*. Whereas individual entities are things such as people, chairs, cities, and rocks, events are things such as walks, dances, discussions, meetings, celebrations, storms, concerts, marathons, goodbyes, births, and deaths.

Events were introduced to semantic theory by the philosopher Donald Davidson at the end of the 1960s to explain the logical properties of predicates. Consider the following pair of sentences:

- (4) Kazimir dances
- (5) Kazimir dances in the garden

Informally, both sentences express the information that Kazimir dances. Sentence (5) provides the further information that Kazimir’s dancing is taking place in the garden. In Chap. 6, we saw that (4) can be understood as attributing a property—that of being dancing—to an individual—Kazimir. The same analysis, however, does not work for (5). In (5), the predicate “dances” does not appear to express a property of individual entities but, rather, a relation between two individual entities—Kazimir and the garden. Hence, a more fitting analysis of (5) would require analyzing “dances” as a relation—as we have done for the verb “admires” in Chap. 7. The conclusion we are led to is that the verb “dances” expresses two different meanings, of different logical types, in the two sentences.

However, having to maintain different semantic analyses for the same verb is undesirable for the following reason. As we can produce sentence (5) from sentence (4) by adding “in the garden”, so we can produce sentence (6) from sentence (5) by adding “at midnight”.

(6) Kazimir dances in the garden at midnight

Should we conclude from (6) that the verb “dance” is actually three-times ambiguous between expressing a property, a binary relation, and a ternary relation? Notice that the same reasoning can, in principle, be replicated *ad libitum*. Sentence (7) can be produced from (6) by adding “with Frida” and sentence (8) from (7) by adding “for Pablo”.

(7) Kazimir dances in the garden at midnight with Frida

(8) Kazimir dances in the garden at midnight with Frida for Pablo

Ultimately, if we were to explain the differences between the various occurrences of the verb “dance” in the different sentences as a form of ambiguity, we would have to conclude that “dance” is infinitely ambiguous.

A further reason for rejecting this approach is that speakers of English share an intuition that the verb “dances” has the same meaning in the different sentences. In all cases, it characterizes Kazimir as performing the same type of action—dancing. The only difference is that in the sentences (5–8) further information is provided concerning where the dancing takes place, when it takes place, with whom it takes place, and for whom it takes place. The intuition is, in other words, that there is a common core to the meaning of all these sentences. There is a formal counterpart to this intuition: Each of the sentences from (5) to (8) logically entails the previous ones. Sentence (5) entails sentence (4): If it is true that Kazimir dances in the garden, then it is also true that Kazimir dances. Sentence (6) entails sentences (5) and (4): If it is true that Kazimir dances in the garden at midnight, then it is also true that Kazimir dances in the garden and it is also true that Kazimir dances. And so on. Sentence (7) entails sentences (6), (5), and (4). Sentence (8) entails sentences (7), (6), (5), and (4). If we were to treat the different occurrences of “dances” in the different sentences as expressing different meanings, we would fail to explain the simple intuition that there is a relation of logical entailment between the different sentences.

Events were introduced by Davidson precisely to solve this problem. According to his analysis, a sentence such as (4) expresses the existence of an event of dancing performed by Kazimir. Slightly more formally, sentence (4) is true if, and only if, there is an event of dancing and the dancer in such event is Kazimir. Compared with the previous analysis, “dances” is now analyzed not as a property of individuals but as a relation between an event and an individual which is true if, and only if, the event is an event of dancing and the individual is the dancer in such event.

This approach solves the problems we reviewed above because now expressions such as “in the garden”, “at midnight”, “with Frida”, and “for Pablo” can be understood as expressing properties of the event of dancing, rather than arguments of the verb “dances”. Sentence (5), for example, can be understood as expressing the fol-

lowing truth-conditions: There is an event of dancing and the dancer in such event is Kazimir and the event takes place in the garden. Clearly, any possible world that verifies these truth-conditions also verifies the truth conditions of (4). Hence, we can explain why (5) logically entails (4).

This brief discussion demonstrates another case in which we find that introducing a qualitatively distinct novel class of objects (in this case, events) is helpful in explaining facts about language (i.e. the entailment relations discussed above) that we could not otherwise (straightforwardly) explain.

## Natural Language Metaphysics Versus Metaphysics Proper

In the decades that followed Montague's first formulation of his theory, semanticists often invoked the need for new essential classes of objects to explain semantic phenomena that resisted a purely quantitative analysis. These include—on top of individual entities, possible worlds, points in time, and events—points in space, situations, facts, tropes, natural kinds, masses, pluralities, and degrees, to mention but a few of the most notable. Generally, semanticists working in Montague's framework motivate the introduction of new essential categories in the models on the basis of strictly *linguistic* considerations. For them, how good a model of linguistic meaning is depends solely and exclusively on how well it explains the facts of language. Take possible worlds. In the previous chapters, we saw that a model that includes possible worlds provides a more satisfactory account of the meaning of declarative sentences than one without them. This is all the evidence linguists need to prefer a model with possible worlds to one without them. Similar considerations motivate the introduction of points in time and events as independent essential classes of objects. For example, a model that includes points in time among its essential elements provides a straightforward account of the meaning of the different verb tenses. A model that does not include points in time amongst its essential elements cannot (arguably) achieve the same result. This is why semanticists prefer a model that includes points in time to one that does not.

Of course, concepts such as times, events, and possibility are also a subject of investigation in the domain of proper philosophical metaphysics. In their practice, however, semanticists maintain a distinction between the metaphysics practiced by philosophers and the metaphysics of natural language they practice. For them, the two disciplines have different purposes. Whereas proper metaphysics is concerned with the question "what sort of things are there?", natural language metaphysics is concerned with the question "what sort of things do people talk *as if* there were?" In effect, it is perfectly possible that what may count as a useful essential categorization when the goal is to provide an account of the nature of reality may turn out to be completely inadequate when the goal is to provide an account of how we talk about such reality. This suggests a fundamental difference between investigating things *as they are*, which is the task of metaphysicians, and investigating things *as we talk about them*, which is the task of semanticists.



This freedom in defining their own metaphysics has allowed semanticists to make enormous progress in explaining a number of phenomena in the domain of natural language meaning. However, the distinction between the metaphysics of things and the metaphysics of meaning also introduces an unwelcome limitation to our quest into the nature of meaning. By keeping natural language metaphysics secluded from the broader understanding of the world, we also close the door to the possibility of understanding how meaning is connected to the world it describes as well as to the individuals who produce it and grasp it. Even when we have identified all of the essential properties that best explain natural language meaning and its interaction with natural language grammar, we are still left with the question of what motivates them. Montague’s approach gives us a glimpse into the essential edges of natural language meaning, its most primitive and essential ingredients. Then, the question we face in our quest towards understanding the nature of linguistic meaning is the following: Why these essential ingredients? What has determined them? What has decided that we should talk about things in this form and not one of the many other, possibly infinite, conceivable ways? What has shaped the essential form of linguistic meaning? It is to these questions that we turn in part II of our book.

## References and Remarks

The term natural language metaphysics is due to Bach (Bach 1986), who proposes a distinction between metaphysics proper—which is concerned with the question “what sort of things are there?”—and natural language metaphysics—which is concerned with the question “what sort of things do people talk *as if* there were?”. An insightful assessment of the issues around natural language metaphysics can be found in recent work by Moltmann (Moltmann 2018).

The first attempt at providing a logical framework for the interpretation of a tensed language is Arthur Prior’s Tense Logic (Prior 1957, 1967, 1969). Prior’s logic is, by and large, at the basis of Montague’s own model of interpretation of Tense in English. The so-called “referential” view of Tense, which has probably become the most popular among practitioners of natural language syntax and semantics, originates with the work of Hans Reichenbach (Reichenbach 1947).

The basis of event semantics is Davidson’s account of adverbial modification (Davidson 1967) according to which the truth-conditions of a sentence such as (i) are expressed by a logical form such as (ii), where “dance” is a relation between an event  $e$  and an individual and both the adverb “beautifully” and the PP “in the garden” express properties of  $e$ :

- (i) Kazimir danced beautifully in the garden
- (ii)  $\exists e[\text{dance}(e, \text{Kazimir}) \wedge \text{beautiful}(e) \wedge \text{in\_the\_garden}(e)]$

The logical form in (ii) roughly reads as “there is an event  $e$  of dancing involving Kazimir and  $e$  is beautiful and  $e$  takes place in the garden”. As we saw, these truth-conditions directly account for inferences among sentences involving adverbial

expressions, for instance, for the inference from (i) to (iii) and (iv) below, discharging the burden of the explanation on the semantics of the Boolean operator “ $\wedge$ ” instead than on specific meaning postulates, as was the case in the approaches to adverbial modification originally inspired by Montague’s calculus.

(iii) Mary danced beautifully

(iv) Mary danced

Davidson’s account was later reinterpreted by Terence Parsons (Parsons 1990) by shifting the reading of action verbal predicates such as “dance” from binary relations to monadic properties of the event argument and explicitly introducing thematic roles as binary relations between the actants of the event and the event itself. In this way, (ii) is reformulated as (v), which roughly reads as “there is an event  $e$  of dancing and Kazimir is the agent in  $e$  and  $e$  is beautiful and  $e$  takes place in the garden”.

(v)  $\exists e[\text{dance}(e) \wedge \text{AGENT}(e, \text{Kazimir}) \wedge \text{beautiful}(e) \wedge \text{in\_the\_garden}(e)]$

Event semantics gave rise to an empirically and theoretically successful stream of research, revealing itself as a stimulating tool for the investigation of complex issues of temporal and aspectual interpretation, and for exploring issues of compositionality at the syntax/semantics interface. However, in spite of its successes, event semantics was also often felt to involve serious (and poorly-understood) compositionality problems. This has led many semanticists to feel uneasy about the event variable in the formalism and to propose alternative frameworks (see, for instance, Beaver and Condoravdi 2007). One such problem is the so-called “event-modification problem”. In event semantics, there is no principled reason why an adverbial modifier such as “with binoculars”, understood as a one-place predicate in a sentence such as (vi), should not apply to the subject argument, giving rise to the interpretation in (vii), besides applying to the event argument, giving rise to the interpretation in (viii):

(vi) John saw a girl with binoculars

(vii) There was an event of seeing a girl, and John, who had or used binoculars, was the experiencer of such an event

(viii) There was an event of seeing a girl, John was the experiencer of such an event, and this event of seeing involved the use of binoculars

We should notice here that whereas (viii) is a legitimate interpretation of (vi), (vii) is not. On the one hand, the interpretation of adverbial modifiers such as “with binoculars” as one-place predicates prompts not only their intersective reading with the verbal predicate, as in (viii), but also a welcome parallelism with the standard cases of (intersective) adjectival modification. On the other hand, once “with binoculars” is interpreted as a one-place predicate, the problem arises which compositionality principle prevents the adverbial modifier from applying, as a property, to the subject argument. This issue (barely discussed in the literature; but see Winter and Zwarts 2011) is compounded by the presence of the perhaps better known event quantification problem. This problem arises in connection with the existential closure requirement in logical forms such as (ii) or (v) above, according to which

the event variable introduced in the logical representation of action sentences must be bound by an existential quantifier in order for the formula to be interpreted as a proposition. This is in itself not a problem (though it raises subtle issues concerning the locus where this operation applies; for example, whether it applies at sentence-level or at a discourse-level), if were not for the fact that the existential operator over events virtually never interacts with the other quantifiers or scopal expressions contained in the sentence. In the case of (ix), for instance, the only possible interpretation of the sentence is provided in (x), whereas the interpretation in (xi), in which the existential quantifier over events takes wide scope with respect to the quantifier expressed by “nobody”, is completely impossible. Once again, we are in need of explaining which compositionality principles bar the event quantifier from taking wide scope, or simply prevent it from interacting with other quantifiers.

(ix) Nobody danced

(x)  $\neg\exists x[\exists e[\text{dance}(e) \wedge \text{AG}(e, x)]]$

“Nobody was the agent of any dancing event”

(xi)  $\exists e[\neg\exists x[\text{dance}(e) \wedge \text{AG}(e, x)]]$

“There was a dancing event where nobody was the agent”

Of course, the event quantification problem raises challenging issues concerning compositionality, prompting an analysis in terms of the presence of distinct compositional layers, or interpretation stages, within sentence-structure, (as, for instance, in the proposals made in Ramchand 2016, inspired by Champollion 2015, according to which the event variable is existentially closed at a very low syntactic level).

## References

- Bach, E. (1986). Natural language metaphysics. In R. B. Marcus, G. J. W. Dorn, & P. Weingartner (Eds.), *Logic, methodology, and philosophy of science, VII* (pp. 573–595). Amsterdam: Elsevier.
- Beaver, D. & Condoravdi, C. (2007). On the logic of verbal modification. In P. Dekker, M. Aloni, & F. Roelofsen (Eds.), *Proceedings of the 16th Amsterdam colloquium*. Amsterdam: University of Amsterdam.
- Champollion, L. (2015). The interaction of compositional semantics and event semantics. *Linguistics and Philosophy*, 38, 31–66.
- Davidson, D. (1967). The logical form of action sentences. In N. Rescher (Ed.), *The logic of decision and action*. Pittsburgh, PA: University of Pittsburgh Press.
- Moltmann, F. (2018). Natural language and its ontology. In A. Goldman & B. McLaughlin (Eds.), *Metaphysics and cognitive science*. Oxford: Oxford University Press.
- Parsons, T. (1990). *Events in the semantics of English: A study in subatomic semantics*. Cambridge, MA: MIT Press.
- Prior, A. N. (1957). *Time and modality*. Oxford: Clarendon Press.
- Prior, A. N. (1967). *Past, present and future*. Oxford: Clarendon Press.
- Prior, A. N. (1969). *Papers on time and tense*. Oxford: Clarendon Press.
- Ramchand, G. (2016). *Situations and syntactic structures*. Cambridge, MA: MIT Press.
- Reichenbach, H. (1947). *Elements of symbolic logic*. New York: McMillan.
- Winter, J. & Zwarts, J. (2011). Event semantics and abstract categorial grammar. *The mathematics of language. 12th biennial conference* (pp. 174–191). Proceedings MOL 12, Nara, Japan, September 6–8.

**Part II**  
**Meaning and Subjects**

## Chapter 15

# Beyond Form



Without doubt, the notion of form—the intangible that remains when all that is meaningful is removed—is one of the pillars of the cultural revolution of the first half of the twentieth century. Malevich—and with him many other artists—used it to liberate art from the chains of the past; Frege, Russell and Wittgenstein to address the fundamental questions of philosophy. It did not take long, however, before the focus on form also disclosed its limitations. Artistic manifestos, such as Malevich's, came soon to be regarded by many as instruments of authority, rather than liberation—norms of conduct determining mechanically the practice of those who endorse it. Similarly, the analytic approach to philosophy initiated by Frege was soon found to be subject to fatal logical flaws. It is at this time that some of the most intricate paradoxes of logic and mathematics are formulated for the first time.

Upon closer reflection, it is not difficult to understand the reason behind these limitations. A formal analysis always begins with a choice. A formal model of a phenomenon is always produced by choosing one or more of its features as essential while disregarding others as non-essential. Consider a simple example. Suppose we want to organize some books on a shelf, which are all by different authors and of different dimension, weight, color, title, genre, year of publication, place of publication, number of pages, and so on. In principle, we could organize our library according to any of these features. In each case, we would achieve a perfectly coherent organization of our library, which elects one feature as the relevant organizing principle and disregards the others as irrelevant. Which feature should we select as the organizing principle of our library is, ultimately, up to us and our needs. Theoretically speaking, however, any option would be as legitimate as the others. The same, one may say in reaction to Malevich's manifesto, goes for images. Images have many different properties—geometric form as well as dimension, material constitution, color, vibrance, dynamicity, negative space, complexity, iconicity, and so on. Each of these properties could be the essential principle at the foundations of a plausible and coherent model of the image. So, why accept Malevich's arbitrary decision that geometric form is the most essential of all?

## World and Mind

In part I, we have seen how Montague's model-theory explains the rich productivity of natural language meaning as the outcome of a restricted number of essential ingredients. What determines these essential ingredients? As the world around us is susceptible to many distinctions, what determines those that are essential to speakers when they talk about it in natural language?

This question has been central to philosophical thought since its very inception. If meaning is what language is about, what is language about? To a large extent, the answers provided can be characterized as belonging to one of two opposing theses. The first is the thesis of *semantic externalism*, which claims that the way we talk about things is determined by the way things are. This thesis owes its name to the fact that it grounds the essential properties of linguistic meaning in the external world of natural objects. The second thesis is the thesis of *semantic internalism*, which claims that the way we talk about things is determined by the way we think about them. This latter thesis owes its name to the fact that it grounds the essential properties of linguistic meaning in the inner psychological life of speakers. Using a different terminology, to which we will return in Chap. 20, we will say that semantic externalism regards meaning as *objective*, whereas semantic internalism regards it as *subjective*.

Whatever your preference at this point, you should know that you are in the company of eminent thinkers. We find the thesis that words refer to mental objects already in Aristotle's *On Interpretation*. It is also the view held by John Locke. The words we utter, Locke writes in his *An Essay Concerning Human Understanding* (III.ii.2), "stand as marks for the ideas in [our] own mind, whereby they might be made known to others, and the thoughts of men's minds be conveyed from one to another." In contrast, John Stuart Mill was a champion of semantic externalism. He famously reacted to Locke's thesis by pointing out that the sentence "the sun causes the day" cannot possibly mean that our idea of the sun causes our idea of the day. What we really mean, rather, is that the sun *as such* causes the day *as such*.

In part II, we will examine both theses in some detail. This exercise will allow us to uncover a number of fundamental features of linguistic meaning. It will also allow us to identify a number of practical and conceptual limitations of both theses. On the basis of these observations, in part III we will explore an alternative view of linguistic meaning, whereby the metaphysical foundations of meaning are to be found at the interface between word and mind—that is, in perception.

# Chapter 16

## Meaning and World



### Semantic Externalism

Semantic externalism is the thesis that linguistic meaning is something that belongs to the external world of material things. The thesis is to be understood in opposition to the thesis of semantic internalism, which is the thesis that meaning is a psychological object, belonging to the realm of the mind. Semantic externalism is also a thesis that has received a notable impulse in the philosophy of language of the twentieth century. In this chapter, we will illustrate two of the most renowned arguments that have been put forward in its favor. Then, in the chapters that will follow, we will consider a number of arguments against it. This exercise will provide us with two main insights. By considering how semantic externalism can be defended against such counter arguments, we will attain a better grasp of its true value. In fact, we will find that meaning, whatever it may be, is inextricably anchored to its material environment. By the same token, however, we will also attain a better understanding of the limitations of semantic externalism. Whatever meaning may be, we will conclude, it inevitably entails an element of subjectivity amongst its essential features, which cannot be reduced to the causal relation that anchors meaning to its material environment.

The first argument we will consider in this chapter is known as the *private language argument* and is due to Ludwig Wittgenstein. The second, more recent, argument is known as the *Twin Earth thought experiment* and is due to the philosopher Hilary Putnam. Both arguments are, as we will see, of a very abstract nature—the sort of philosophical argument that, to borrow the words of the philosopher Mark Rowlands, “make non-philosophers despair of philosophers”. Yet, it is essential that we understand them in some details.

## Private Language (and Why You Shouldn't Try It)

Imagine yourself experiencing a certain sensation, such as some type of subtle pain. You decide to register the occurrence of this sensation by marking your diary with a symbol, say, the capital letter "S". The symbol registers the occurrence of the sensation on the relevant day. What you have created is quite unique—in fact, an extraordinary puzzle to philosophers and logicians. You have created a *private language*, that is, a language that can only be understood by a single person, its creator.

It is important to appreciate that the language you have created is private for reasons than are not just accidental. To understand why, it is useful to draw a distinction between two types of private languages: *contingently* private languages and *logically* private languages. Contingently private languages are languages that are private for contingent, accidental reasons. These are languages that are in principle understandable by more than one person, but, in the circumstances in which they happen to be employed, are known only to a single individual. Logically private languages, conversely, are languages that cannot be understood by anyone but their creator, not even if the creator were to make the conventions of the language available to others.

It is easy to think of examples of contingently private languages. Consider the alternative scenario in which you decide to mark your diary with the capital letter "T" every time your telephone rings. Such a linguistic convention could in principle be learned and understood by others. Still, you can decide to keep it secret. If you do so, your language does qualify as private but only in a contingent sense. An example of a logically private language is the language you have imagined above. It is necessarily private because its only symbol, "S", refers to a sensation. Sensations are private by their very nature, as they can only be known to the individuals who experience them. You, as the creator of the language and experiencer of the sensation are the only one who can set and understand the conventions of your language. No one else but you can grasp the meaning of "S" or employ the symbol with the meaning you intend it to have.

Why is this language so puzzling to philosopher and logicians, then? In his last work *Philosophical Investigations*, Ludwig Wittgenstein formulated an argument that demonstrates that a logically private language is, in fact, logically impossible. To see why, return to the private language you have created. You have so far established a symbol "S" to refer to a sensation you have experienced. Suppose that the day after you have introduced the symbol "S", you experience a second sensation, also of pain. This pain is very similar to the pain you have experienced the day before, although, perhaps, not completely identical. Should you note the sensation as "S" or introduce a new symbol for it? Whatever your inclination at this point, you should first notice that the question is a tricky one. To begin with, you should consider that there really are only two ways to go about it. Either you decide to note the new sensation with a different symbol, therefore adding a new symbol to the vocabulary of your language, or you decide to note it as "S", in which case you are altering the existing rule of interpretation of the symbol "S".



Consider the two options in some more detail. Consider, first, what happens if you decide that today's sensation is different from the one you experienced yesterday and introduce a new symbol to note it in your diary—say, the capital letter “P”. By doing so, you have effectively introduced a new symbol in the lexicon of your language and a new convention concerning its interpretation. What you have done is not merely employing the language you had created yesterday. Rather, you have extended the language by, in effect, coming out with a new word for a new object.

Suppose, alternatively, that you decide that today's sensation is similar enough to the one you experienced yesterday and, accordingly, resolve to annotate it with the symbol “S”. In this case, you are not introducing a new symbol but you are still establishing a new rule of interpretation. You are rewriting your current rule of interpretation of the symbol “S”, which you originally introduced to refer to yesterday's sensation, with a new rule whereby “S” is now allowed to refer to yesterday's sensation as well as today's. You are, that is, taking an existing symbol to mean today something more than what it meant yesterday.

Ultimately, your choice is going to require either the introduction of a brand-new rule of interpretation or the reformulation of an existing one. Furthermore, if you proceeded with the experiment, you would find that, day after day, sensation after sensation, the problem would arise at every occasion of use of the private language. Not once, you would be able to use your private language by simply employing an already existing rule of the language. On each occasion, you would have to take one of the two decisions above—that is, either introduce a new symbol or revising the rule of interpretation of an existing one. Either way, it seems a private language is something that is never merely used but is always and necessarily undergoing a process of constant re-creation.

There is a further reason for preoccupation. Whatever your answer to the dilemma of whether to use an existing symbol or introducing a new one, it is bound to be right. To be more precise, there really is no right or wrong answer. The dilemma merely puts you in front of a choice that you, the creator and owner of the language, are free to take in the privacy of your mind. It's up to you to decide how to characterize the sensation you feel and either decision you make is going to be just correct for as long as you decide it to be so.

If this is true, however, it voids the language of any real value, as Wittgenstein warns us. In effect, the language you have created cannot be described as the product of a set of pre-established rules—the result of the systematic unfolding of a limited set of principles that are set a priori. It is, as we have observed, a constant process of creation. Henceforth, whatever the rules of interpretation of the language, they do not translate into corresponding *rules of use*—rules, that is, that allow the user of the language to predict and discriminate the correct uses of a symbol from the incorrect ones at any future occasion of use of the language. Whether a symbol is used correctly or not does not depend on the rules of the language. Rather, it is something the user of the language decides arbitrarily at each occasion of use of the language.

To better appreciate this point, we can use an analogy between the rules of a language and the rules of a game. Let us take, for example, the popular board game

“Game of the Goose”. Players move in turns along a track by throwing a dice. Each space on the track is numbered, from 1 to 63, and is associated with a specific instruction. A player who lands on 6, for example, must move forward to 12. A player who lands on 42, must return to 30. And so on. The winner is the first player who reaches 63. Consider now the following variation of the game. The rules remain the same but, this time, whenever a player lands on a new space on the track, it is up to her, and only to her, to decide which number should correspond to that space. It is easy to see that, in such a game, it hardly makes any sense to talk of rules. What a player does is, in effect, to freely decide at each step of the game what the outcome of its move is. The player may very well decide, after throwing the dice for the first time, that the space where she landed is number 63 and, in this way, win the game. Needless to say, this version of the game renders the game dull. But so then is a logically private language. As in the modified version of the Game of the Goose, it is up to the creator of the private language to decide, at each occasion of use of a symbol of the language, whether the use of a symbol is correct or incorrect in its application to the current circumstances.

The reason why a private language is bound to run into this problem is obvious: Its model of interpretation is the inner psychological life of its one and only user. In effect, the one and only user of a private language is in charge of deciding, at once, the rules of the language as well as the circumstances against which these rules are evaluated. This, however, renders the language as dull as the modified version of the Game of the Goose.

Since its first formulation, Wittgenstein’s argument against private language has been the subject of heated philosophical debates. From our perspective, the general conclusion we should draw is that, if the meaning of a language is what language is about, then meaning cannot be something that belongs to the private minds of speakers, because that would render the language as hollow and dull as a game with no rules.

## **The Twin Earth Thought Experiment**

The second argument we will consider in support of the thesis of semantic externalism is the so-called “Twin Earth thought experiment”, by the philosopher Hilary Putnam. The first thing we should observe about Putnam’s argument is that it is a “thought experiment”. A thought experiment is a form of argument used—mostly by philosophers, but not only—to support a conclusion by comparing two imaginary state of affairs. It is called “experiment” because it is based on the same methodological principles governing actual scientific experiments. That is, it is a procedure for identifying the causes of a phenomenon by observing it in two contrasting environments. Suppose you are studying a certain phenomenon, call it *P*, and wish to identify its cause, that is, the factors that uniquely and exclusively bring *P* into being. Suppose, further, that you have formulated the hypothesis that a certain factor *C* is the cause of *P*. To test your hypothesis experimentally, you can

compare two states of affairs that are in all respects identical except in that one includes *C* and the other does not. If *P* materializes in the state of affairs that includes *C*, but not in the one that excludes it, then your hypothesis is proven as correct. A thought experiment qualifies as an experiment because it follows the same procedure. What is peculiar about it is that, instead of comparing two actual states of affairs, it compares two imaginary ones. This, however, is not a reason to underestimate the importance of thought experiments. Among other things, they are part of our daily lives as when, for example, we compare two different hypothetical courses of action before making an important decision.

Putnam's Twin Earth thought experiment compares two imaginary scenarios in order to demonstrate that the meaning of linguistic expressions depends on factors that are independent of the minds of the speakers who use such expressions. The first state of affairs in the comparison is the actual planet Earth, with all of its constituents, natural elements, history, and the life forms that populate it. The second state of affairs is an almost perfect duplicate of planet Earth—a Twin Earth that is identical to actual Earth in almost every single respect. For each individual, animal, plant, and rock on actual Earth, there is an exact duplicate on Twin Earth. The two planets share the same history, as do all of their constituents, including the individuals that inhabit it. In fact, on Twin Earth there is also an exact duplicate of you, with the same physical constitution, history, behavior, thoughts, beliefs, and dispositions. Everything you have done, thought, and believed, your duplicate has also done, thought, and believed. Twin Earth differs from actual Earth only in one respect. Whereas water has the chemical composition  $H_2O$  on actual Earth, it has a different chemical composition on Twin Earth. That is, also on Twin Earth there is a liquid, colorless, and transparent natural substance that runs in rivers, fills lakes and oceans, and is sold in bottles; people use it to wash and refresh themselves and, unless one runs a very detailed chemical analysis, it is undistinguishable from the water we find on actual Earth; furthermore, English speakers on Twin Earth call this substance "water," as they do on actual Earth. Yet, this natural substance is, ultimately, of a different sort than the water we find on actual Earth.

Consider now an individual, call him actual Kazimir, who is a speaker of English on actual Earth. As Twin Earth is identical to actual Earth in all respects, except in the constitution of water, the very same individual lives also on Twin Earth. Let us call him, Twin Kazimir. Actual Kazimir and Twin Kazimir are in all respects identical. They are made of the same physical elements. They share the same history, thoughts, beliefs, and dispositions. Suppose now that the two Kazimir's utter—of course, simultaneously—the word "water" and consider the following question: Does the word "water" uttered by actual Kazimir on actual Earth mean the same thing as the word "water" uttered by Twin Kazimir on Twin Earth?

Again, this is a difficult question. Of course, it may well be the case that actual Kazimir and Twin Kazimir are able to fulfill the same intentions and achieve the same goals by uttering the word water. If they were to ask for a glass of water to refresh themselves, they would be able to achieve the same goal by uttering the same words. Yet, Putnam warns us, the word "water" itself, considered independently of the intentions of its users, means, in effect, different things on the two

different planets. As actual water and twin water are two different substances, then the word “water” means different things in the different planets. In fact, if we were to visit Twin Earth and study its water, we would conclude that, after all, “it is not water”!

The point of the thought experiment is that, if we agree with Putnam that the word “water” means different things in the mouths of the two Kazimir’s, we must conclude that the meaning of the word “water” is a function of the material environment in which the word is used and crucially does not depend in any way on what goes on in the minds of the speakers that use it. What “water” means depends on what water is and not on what the two Kazimir’s think it is because, in the experiment, the two Kazimir’s are completely identical to each other, down to the smallest physical units. If they have identical brains, then they must also have identical thoughts when simultaneously uttering the word “water”. The fact that the word “water” can have different meanings even when uttered by speakers with completely equivalent thoughts, demonstrates that what “water” means depends on what water is and not on what speakers think it is. As Putnam himself puts it, “if you fix everything that is going on in the head, and vary the environment, then meanings [...] will vary with the changes in the environment even though nothing has changed inside the head.”

## Meaning Ain’t in the Head

Wittgenstein’s private language argument and Putnam’s Twin Earth experiment point, in different ways, to the same conclusion, which is most aptly summarized by Putnam’s famous slogan, “Cut the pie any way you like, ‘meanings’ just ain’t in the head!” Meaning, that is, is not a mental object. The factors that determine the meaning of a linguistic expression must be found in the external world of material objects not in the inner realm of the mind. This is the thesis of semantic externalism.

## References and Remarks

Rowland’s quote is from his 2003 book *Externalism* (Rowlands 2003, p. 103), which is a critical essay on the thesis of externalism, although not merely conceived as a theory of linguistic meaning but also as a more general approach to the nature of mental contents and states. Among other things, the book offers excellent introductions to the most significant philosophical arguments in favor of externalism. These include Wittgenstein’s private language argument and Putnam’s Twin Earth thought experiment, which we reviewed in this chapter, but also other arguments that had an important role in setting up the externalist framework and its agenda—such as Tyler Burge’s thought experiment of Counterfactual Earth (Burge 1979) and

David Kaplan's logic of indexicals (Kaplan 1989a, 1989b). We will return to indexicals and Kaplan's logical framework in Chaps. 21 and 22.

Wittgenstein's private language argument was originally formulated by Wittgenstein in his last work *Philosophical Investigations*. The work was published for the first time in 1953, two years after Wittgenstein's death, by G. E. M. Anscombe (Wittgenstein and Anscombe 1953). It offers an extraordinary testimony of Wittgenstein's late philosophy, not only because of its content, but also because of its unique style of exposition. Notably, the private language argument is not developed in an orderly, linear fashion, with clearly defined premises and conclusions. It is rather suggested by various considerations made at different points in the text. Even the label "private language argument" is not due to Wittgenstein himself, but to later commentators of his work. The argument, as it is most commonly understood in current philosophical discussions, is the result of a long process of reconstruction and interpretation performed by a number of different scholars. Different interpretations of Wittgenstein's original insights have been proposed along the years, sometimes leading to very different conclusions. What the deeper implications of the private language argument are remains still a very disputed issue. The literature on the topic is vast. A canonical reference is Kripke's interpretation of Wittgenstein's private language argument (Kripke 1982), which will be discussed in the next chapter. For those interested in approaching Wittgenstein's original text we recommend Marie McGinn's guidebook (McGinn 2013).

Putnam formulated his Twin Earth thought experiment in his 1975 article "The meaning of meaning" (Putnam 1975). The quote at the end of the chapter is from this article. Putnam's argument was originally formulated as an argument concerning linguistic content and was only later generalized to mental contents by Colin McGinn (McGinn 1977). Putnam's original article includes other arguments in favor of semantic externalism and other arguments were presented in his subsequent work (for example, in Putnam 1983). The Twin Earth thought experiment is certainly the most well-known amongst Putnam's arguments. This despite the fact that some of the other arguments are more realistic and less far-fetched. A good example (from Putnam 1983) concerns the word "grug", which is used in the imaginary country of Ruritania to refer to the metal used to make pots and pans. The linguistic community of Ruritania is divided in two different variants. In South Ruritania "grug" refers to silver whereas in North Ruritania "grug" refers to aluminum. The question Putnam raises is then whether the sentence "Pots and pans are made of grug" has the same meaning when uttered by speakers from the north and from the south of the country. The answer, according to Putnam, is that the two sentences have different meanings because the word "grug" picks up different referents in the two dialects. The conclusion is, once again, that "Meanings just ain't in the head". Even in the case in which speakers of North and South Ruritania have identical mental representations or conceptions of "grug", they still refer to different objects when using that word and, therefore, convey different meanings.

Putnam's arguments have been reviewed and criticized in a number of ways by a number of authors. We will discuss some of this criticism in Chap. 30. For the time being, it is worth mentioning at least Fodor's criticism of Putnam's argument, as

formulated in his book *Psychosemantics* (Fodor 1987). Fodor's criticism is especially important because it relies on a distinction between two types of content—a *broad* content, corresponding to the meaning an expression has in its environment of use, and a *narrow* content, corresponding to the meaning an expression has in the mind of its users, irrespectively of its relation to the environment. Once this distinction is drawn, it is possible to reconcile the fact that expressions such as “water” or “grug” take different referents when used in different environments with the fact that they have the same cognitive significance to their users. As we will see at the end of Chap. 30, the difficulty of this approach resides in providing a satisfactory definition of narrow content.

## References

- Burge, T. (1979). Individualism and the mental. In P. French, T. Uehling, & H. Wettstein (Eds.), *Midwest studies in philosophy* (Vol. Vol. 4, pp. 73–121). Minneapolis, MN: University of Minnesota Press.
- Fodor, J. A. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.
- Kaplan, D. (1989a). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). Oxford: Oxford University Press.
- Kaplan, D. (1989b). Afterthoughts. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 565–614). Oxford: Oxford University Press.
- Kripke, S. A. (1982). *Wittgenstein on rules and private language*. Harvard: Harvard University Press.
- McGinn, C. (1977). Charity, interpretation, and belief. *Journal of Philosophy*, 74, 521–535.
- McGinn, M. (2013). *The Routledge guidebook to Wittgenstein's philosophical investigations*. Abingdon: Routledge.
- Putnam, H. (1975). The meaning of ‘meaning’. *Minnesota Studies in the Philosophy of Science*, 7, 131–193.
- Putnam, H. (1983). Computational psychology and interpretation theory. In D. M. Rosenthal (Ed.), *The nature of mind* (pp. 527–535). New York: Oxford University Press.
- Rowlands, M. (2003). *Externalism*. Chesham: Acumen.
- Wittgenstein, L., & Anscombe, G. E. M. (1953). *Philosophical investigations*. London: Basic Blackwell.

# Chapter 17

## Meaning and Actions



### Meaning and Rules

Starting from this chapter, we will consider a number of objections against semantic externalism. By reviewing these objections, we will be able to refine our understanding of the conceptual pillars at the foundation of the thesis as well as identify some of its weaknesses.

The first objection we will consider concerns Wittgenstein's private language argument. Wittgenstein shows us that the rules of interpretation of a logically private language cannot translate into rules of use. The rules of a private language are *indeterminate*, in the sense that we cannot rely on them to predict the validity of its future applications. Against this conclusion, in this chapter we will consider the objection that, in practice, the logic that applies to private languages also applies, with similar disruptive consequences, to public ones. At a closer look, in fact, we find that the rules of interpretation of expressions referring to public objects also fail to translate into fully deterministic rules of use.

As an example, consider the common noun "table". Assume, following our current theory, that its meaning is defined as the set of objects that are tables—or, if you wish, its corresponding characteristic function. At a first look, this definition appears to easily translate into a corresponding rule of use—a rule, that is, that allows its users to judge as correct or wrong any future use of the word. We can formulate it as follows: The word "table" is used correctly whenever uttered in reference to objects that are tables, incorrectly otherwise. This rule tells us what counts as a correct use of the word and what does not. If we say "table" while referring to a table, we are using the word correctly. If we say "table" while referring to something that is not a table, we are using it incorrectly. This seems clear enough.

We should observe, however, that the ability to use such a rule depends, in practice, on the ability to tell whether something is or is not a table. To determine whether a given use of the word "table" qualifies as correct, that is, we must be able



to determine whether the object the word is used in reference to qualifies as a table. Do speakers of English possess such ability? That is, do they possess a well-defined criterion for discriminating objects that are tables from objects that are not? The fact that speakers of English appear to be able to use the word “table” effectively suggests that they must possess some criterion for distinguishing tables from non-tables. This also seems obvious enough. Yet, it is easy to show that if speakers of English possess a criterion of this sort, this criterion is not a deterministic one.

Consider, as a counterexample, the art piece called “Table with two legs on the wall”, first presented by the contemporary Chinese artist Ai Weiwei in 1996. This art piece consists of a table that has been modified so that two of its legs are oriented horizontally and placed against a vertical wall. The result is a table that functions, in effect, as a chair. Let us ask ourselves: Does or does not Ai Weiwei’s creation qualify as a table? In favor of a positive answer, we may observe that the object has four legs and a tabletop, which is a prototypical feature of tables. In favor of a negative answer, we may also observe that the object is really made so that you can sit on it, which is the characteristic of a chair. Should we then refer to this object as a “table” or as a “chair”? What is interesting about this question is that it is strikingly similar to that raised by Wittgenstein in his private language argument. Whether the object deserves or does not deserve to be called a “table” ultimately depends on an arbitrary decision on the part of the users of the language. The use of “table” in reference to the new object cannot be judged by merely applying an existing rule of interpretation, a new one must be defined or an existing one must be redefined.

The circumstances we have now exemplified with the common noun “table” can be reproduced with potentially all expressions of English and, in fact, any public language. They occur every time we encounter an object to which the existing rules fail to apply successfully. Does Pluto deserve to be called a “planet”? Does Duchamp’s urinal deserve to be called “art”? Does 4’33”—John Cage’s famous composition consisting of four minutes and thirty-three seconds of absolute silence—deserve to be called “music”? The answers to these questions ultimately depend on what speakers decide counts as a planet, music, or art. And that is a matter of stipulation in the same way as it is a matter of stipulation whether today’s sensation of pain qualifies as similar enough to the one experienced yesterday. In the end, the rules of interpretation of a public language appear to be as indeterminate as those of private language and that is because it is always possible to find a new object of reference for which the existing rules of interpretation fail to apply straightforwardly.

But then, if public language is as inconsistent as private language in constraining its use, should we conclude that public language is also a dull, hollow exercise? It is easy to find this conclusion unacceptable. For one thing, it contradicts the mere observation that public languages are alive and well and effectively function as means of communication. If we find the conclusion that public languages are as impossible as private ones unacceptable, then we must also conclude that there is something wrong with Wittgenstein’s argument.



## Meaning as Use

According to a notable interpretation of Wittgenstein's argument, due to Saul Kripke, the indeterminacy of the rules of public language is not really an objection to the private language argument but, rather, the broader conclusion that we should draw from it. According to this interpretation, the ultimate target of Wittgenstein's argument is not just private language, but rather the very notion of a rule of interpretation. Private language simply offers the best case to demonstrate that a language is not the deterministic outcome of a set of pre-established rules.

The notion of a rule of interpretation is at the core of the theory of meaning we have discussed in part I. The meaning of simple expressions, as we saw, is set in advance—that is, established a priori—by rules of interpretation associating the simple expressions with the objects they refer to. The meaning of complex expressions is also provided by a general rule of interpretation—the rule of functional application—which combines mechanically the meanings of simple expressions according to the grammatical structure that holds them together. According to the interpretation of Wittgenstein's argument we are now considering, it is a mistake to explain the functioning of a language in this way—that is, as the unfolding of a closed set of pre-established linguistic rules. The reason is that linguistic rules cannot translate into corresponding rules of use and, henceforth, they cannot explain the linguistic behavior of speakers. The way speakers use language should just not be understood as the automatic product of some a priori knowledge they have of its rules.

Yet, if language is not the product of linguistic rules, what is it? At several points in the *Investigations*, Wittgenstein suggests an answer to this question that requires replacing the view that language is made of rules with the view that language is made of *actions*. According to this suggestion, we should not look for meaning in the principles that govern the reference of linguistic expressions but, rather, in the *use* that is made of them. Instead of looking for the a priori principles that determine the form of the language, we should observe the a posteriori effects language produces on its environment in the here and now of its employment. Paragraph §2 of the *Investigations* is the earliest and, perhaps, most notable occasion in which Wittgenstein contemplates this view. Here, Wittgenstein introduces the notion of a *language-game* by presenting the example of a “primitive” system of communication between a builder and his assistant:

The language is meant to serve for communication between a builder A and an assistant B. A is building with building-stones (Bausteine): there are blocks, pillars, slabs and beams. B has to pass the stones, and that in the order in which A needs them. For this purpose they use a language consisting of the words “block”, “pillar”, “slab”, and “beam”. A calls them out; – B brings the stone he has learnt to bring at such-and-such a call.

In the system of communication described in this passage, the expressions “block”, “pillar”, “slab”, and “beam”, have meaning only in as much as they serve their function in the construction enterprise. What really matters is not what the words of the language refer to but, rather, that they produce the desired effect—that

is, that the assistant fetches a block whenever the builder says “block”, a pillar when he says “pillar”, and so on.

The view of language that Wittgenstein suggests replaces the notion of reference with the notion of use. Linguistic expressions do not refer to things, they rather *perform a function*. For this reason, the view is often referred to by as “theory of meaning as use”. This view successfully addresses the objection we have raised at the beginning of the chapter. Once the focus is moved from reference to use, the issue of translating linguistic definitions into rules of use simply dissolves because there are no longer linguistic definitions to begin with. Whether an expression is used correctly or not depends solely and exclusively on how well it performs its function.

## Speech Acts

There is an important lesson we must learn from Wittgenstein: Speaking a language is part of an activity. We do things with words: We make statements, give orders and instructions, describe things, report events, speculate, formulate hypotheses, make up stories, act plays, sing catches, tell jokes, solve problems, ask, thank, curse, greet, play, swear, and promise. These are the actions we perform with language, for which the philosopher John L. Austin introduced the term *speech acts*.

There are a number of examples of expressions whose meaning is determined by the function they performed. Consider the question in (1).

(1) Can you pass me the salt?

Literally speaking, (1) asks whether you—the addressee—are in the capacity of passing me—the speaker—the salt. But this is not what (1) really means. Any proficient speaker of English would interpret (1) as a request to have the salt. When asking (1), we do not want our addressee to say, “yes, I can pass you the salt”. We just want the salt. The meaning of (1) has obviously more to do with the effect the statement produces than with what it is about.

Another example is that of exclamations such as “water!” The value of yelling “water!” is, again, determined by the function that is performed by doing so—for example, indicating that water has finally been found, that a flood is approaching, or that the roof is leaking. Also in this case, the semantic value of the linguistic expression relies in the effect it elicits in the actual circumstances of its use.

Wittgenstein’s observations have historical importance because they initiated a new stream of inquiry in philosophy focused on the actionality of language. Champions of this approach are philosophers such as Austin—whom we already mentioned above –, Paul Grice, John Searle, and Peter Frederick Strawson. Their work has contributed a number of fundamental insights as well as the development of explicit theoretical frameworks—including some formal ones—that capture the complex interaction between language, meaning, and actionality. Today, the actionality of language is fully recognized as one of the determining factors of linguistic

meaning and is accounted for by theories which, in many cases, are fully compatible with the formal approach to language we have explored in part I.

## Meaning as Use Means No Meaning

Despite this important lesson, there are also important objections to the theory of meaning as use. To begin with, the theory of meaning as use qualifies as a theory of *meaning as nothing*, in the sense we discussed in Chap. 3. The theory of meaning as use does not simply claim that use is one of the factors that contribute to meaning but, rather, that use is the *only* factor that determines meaning. What we call meaning is but the emergent product of the relations that hold between linguistic expressions and the effects they produce on the environment in which they are employed. As such, the theory of meaning as use is affected by the same problems that generally affect all theories of meaning as nothing. Whereas it is clear, as we have discussed above, that the function performed by an expression may contribute to its meaning, it is also clear that use is not the only factor that is relevant to meaning. On the contrary, it is the fact that expressions have a meaning of their own that enables speakers to put such expressions into practical use. The expressions we discussed above—such as the question “could you pass me the salt?” or the exclamation “water!”—enable speakers to achieve their goals precisely because they have a literal meaning of their own. It is because the question “could you pass me the salt?” means what it means that speakers are able to employ it as a request to pass the salt. Similarly, it is because the word “water” means what it means that speakers are able to use it and interpret it as, say, a warning that the roof is leaking.

As the structuralist view of meaning failed to explain meaning on the basis of grammatical relations alone, so the theory of meaning as use fails to explain meaning on the basis of use alone.

## Meaning and Linguistic Theory

There is a more general objection to the theory of meaning as use. The theory sets its foundations on the negative claim that the facts of language *cannot* be explained as the unfolding of an underlying system of pre-established rules and definitions. There is nothing a priori to language. Language only exists in the here and now of its use and function. Admitting that there is no a priori system of principles at the foundations of language, however, amounts to admitting that language is something we cannot describe in the rigorous terms of a scientific theory. To provide a theoretical account of a phenomenon is, after all, to explain it as the outcome of general principles. If we agree that language is not—and cannot be—the outcome of some general principles, we must also conclude that language is not something of which we can provide a rigorous theoretical account.

If this is correct, the theory of meaning as use is a somewhat contradictory creature. It is a theory of meaning whose central claim is that there cannot really be a theory of meaning. According to some commentators, this contradiction arises from a mistaken interpretation of Wittgenstein. His intent, they claim, was never that of providing a theory of meaning, but, in fact, that of demonstrating its impossibility. According to these commentators, the lesson we should learn from Wittgenstein is to stop trying to explain the facts of language and, instead, observe them as they occur spontaneously in their environment. “Don’t think, look!”, as Wittgenstein himself famously put it.

To the modern linguist, however, this conclusion is hardly acceptable. Linguists found a way to study natural language with scientific rigor precisely by looking at it as a system of general principles. Saussure was the first to understand that the meaning of any linguistic sign is stipulative and, therefore, arbitrary. Chomsky was the first to view language as the result of a generative system of predetermined rules and definitions capable of compensating for its arbitrariness and, at the same time, explaining its productivity. Montague’s theory of meaning rests on the same conceptual pillars. These theoretical attempts have allowed linguists to provide predictive accounts of a number of different aspects of natural language and, more often than not, to test their predictions experimentally.

## Language and Learning

A last, more practical objection to the theory of meaning as use concerns language learning. The view of language suggested by Wittgenstein brings with it a precise view of the process of learning a language. If the value of language resides in the a posteriori effect it produces on its environment of use, then language must be mastered through a procedure of *trial and error*. Learners attempt different linguistic actions and identify which ones to maintain and which ones to discard on the basis of the feedback they receive from the environment. Imagine a child being born in a linguistic community—a community of speakers sharing a common language. During her growth, the child hears the speakers around her making sounds with their mouths and, perhaps guided by an innate instinct to imitate their actions, starts producing sounds of her own. These sounds trigger different effects on her environment and reactions from the other speakers around her—some positives, some negative. As a result of this feedback, the child refines her language skills and, eventually, achieves the same degree of proficiency as the other members of the community. This is a rather intuitive view of language learning. For one thing, that language is learned as a function of the environment seems quite obvious considering that people learn different languages in different environments. Furthermore, learning by trial and error is a mode of learning we are familiar with in many other domains of human cognition.

This view of language learning was certainly the most widely endorsed around the time of Wittgenstein's *Philosophical Investigations*. It is, in fact, the view expounded by the behavioral psychologist Burrhus Frederic Skinner in his book *Verbal Behavior*, published in 1957. This view underwent fierce criticism in the decades that followed, starting from Chomsky's 1959 review of Skinner's book. Chomsky was the first to point out that feedback plays only a marginal role in language learning. We know today from a number of studies on the interaction between children and adults that the feedback children receive from their environment when mastering their native language is very limited and almost exclusively positive. Only very seldom children receive explicit correction of their linguistic behavior from the adults and, when they receive it, it is generally ineffective. Yet, despite the lack of useful feedback, children manage to master their native language quite naturally, at great speed, and with great consistency across different languages, cultures, and parenting styles. In reality, language learning is not a straightforward function of the learner's environment, as predicted by the view suggested by Skinner. Children do not learn language by first making a lot of errors and then improving on them thanks to the feedback they receive from their environment. To the contrary, language is something that grows with the child's cognition. Following a path that begins with their birth, young learners go on increasing their linguistic repertoire steadily and resolutely until puberty when, unless hindered by specific neuropsychological conditions, they reach full grasp of their native language.

These considerations have been taken to suggest that children are guided in the process of learning their native language by an innate body of cognitive principles that allow them to swiftly and effectively translate the scattered evidence from their environment into the full complexity of a natural language. These principles are an intrinsic component of human nature and explain why different children learning different languages in different cultural environment are instinctually driven to learn language in the same way and with equivalent results.

As of today, the debate about what is provided by the environment and what by the learner's cognitive endowment in language learning is still very open. Yet, only very few would disagree that both elements are necessary. Whereas it remains obvious that language is learned from the environment, it is also true that learners approach this task by relying on a cognitive machinery that is genetically predisposed for the task.

Against the predictions of the theory of language as use, the complexity of the process of mastering a language cannot be reduced to the sole interaction between speakers and their environment. There is something more fundamental and systematic that constrains the process—a body of knowledge that guides the linguistic predictions of the learner. This observation, supported today by an important body of experimental evidence, contradicts the thesis that language exhausts its value in the here and now of its use and the feedback it elicits from its environment.

## Meaning and Community

Before we conclude this chapter, we should observe that there is yet another strategy—alternative to the theory of meaning as use—to respond to the objection that public language rules are exposed to the same indeterminacy as private ones. This strategy capitalizes on the fact that the rules of public language are, in fact, *public*—that is, they are the domain of a community of individuals, rather than that of a single, isolated individual. For this reason, this approach is known as the *community view*.

To introduce it, let us briefly return to the parallel with the Game of the Goose, which we used in the previous chapter to illustrate the problem raised by the notion of a logically private language. There, we compared logically private languages to a modified version of the Game of the Goose, where its unique player is in charge of both throwing the dice and assigning numbers to the cells on the board. As we saw, this peculiar setting renders the rules of the game dull, because they basically allow the player to decide her moves arbitrarily. It is the player herself who, by deciding the numbers on the cells where she lands, assumes unbound control of the effects of all her actions.

The objection to the private language argument we are considering in this chapter is that the rules of public languages are, after all, as arbitrary as those of private ones. In the terms of our parallel with the Game of the Goose, the objection could be formulated as follows. A public language is not so different from a private one. It is like the modified version of the Game of the Goose in that its players are in charge of deciding which numbers go on which cells. The only difference is that, this time, there are not one but many players.

To the supporters of the community view, however, this is a critical difference. Whereas the community view admits that the rules of public languages are vitiated by a degree of indeterminacy, it also points out that, in the context of a public language, this indeterminacy is resolved through agreement within the larger community of speakers who share the language. To return to our metaphor, the community view regards public language as like the modified version of the Game of the Goose with the provision that the numbers on the cells must be agreed by all the participants in the game. This difference is a substantial one as it returns value to the rules of the game. Now, it is not up to each single individual to decide which moves are legitimate and which are not. It is rather something that is negotiated and decided by the community of players at large. Public language rules, precisely because they are public, are subject to the negotiations, disputes, and settlements of all social contracts. Whereas it is true that the decision of what counts as a table is, after all, the result of a mere stipulation, it is also true that such stipulation must be accepted by the community of speakers at large. Whether the linguistic behavior of an individual qualifies as correct depends, ultimately, on how that behavior conforms to the practices, customs, and expectations of the larger community of language users.

The community view, however, also has its shortcomings. To begin with, according to the community view the fact that language is a public business is a necessary

condition for the very existence of language. If using language requires the capacity of following rules, but following rules requires being a member of a community, then language must be public. There is, in other words, no way a language can exist outside the domain of a community of more than one speaker. So formulated, the thesis not only excludes the possibility of logically private languages, but also the possibility of contingently private ones—that is, languages that are private only for contingent reasons. This conclusion is, according to many, too strong because it completely excludes the possibility that an isolated individual be able to develop a language.

Another, more general, objection to the community view is that it ultimately faces the same problems of the theory of meaning as use. In fact, as the theory of meaning as use, the community view attempts at reducing language and meaning to the sole interaction between speakers and their environment. The only significant difference is that the community view identifies the environment with the linguistic community and its social conventions. Otherwise, the community view is subject to the same criticisms we addressed towards the theory of meaning as use.

## References and Remarks

The view that the meaning of linguistic expressions corresponds to their use is commonly traced back to paragraph §43 of Wittgenstein's *Philosophical Investigations*: "For a *large* class of cases of the employment of the word 'meaning'—though not for all—this word can be explained in this way: the meaning of a word is its use in the language" (Wittgenstein & Anscombe 1953).

A behaviorist understanding of language was also promoted by Quine. As we mentioned in the references and remarks section at the end of Chap. 3, he developed arguments, such as the problem of radical translation, to support the view that language should be studied as a form of behavior, rather than as a system of rules. We refer the interested reader to the relevant section in Chap. 3 for more information and references.

Kripke's interpretation of the private language argument is illustrated in his essay *Wittgenstein on Rules and Private Language* (Kripke 1982). We refer the reader interested in the community interpretation of Wittgenstein's remarks to John V. Canfield's article "The community view" (Canfield 1996).

The field of linguistics that concerns itself with the relation between meaning and actionality is called *pragmatics*. Pivotal to the development of pragmatics as an independent field in linguistics were the contributions of scholars such as Austin, Grice, Searle, and Strawson, to mention only a few (see Austin 1962; Grice 1975; Searle 1969; Strawson 1964). An important aspect of pragmatics concerns those expressions whose semantic value is a function of the circumstances in which they are employed. Expressions of this sort include *indexicals*—such as "I", "you", "here", and "now"—and *demonstratives*—such as "there", "this", and "that". Indexicals



will be the topic of Chaps. 21 and 22. Today, there are many textbooks available to those interested in an introduction to the field and its many dimensions.

Skinner's 1957 essay *Verbal Behavior* (Skinner 1957) was criticized by Chomsky in his 1959 review of Skinner's book (Chomsky 1959). A less known essay by Chomsky on the topic of this chapter is "Some empirical assumptions in modern philosophy of language" (Chomsky 1969), which provides relevant criticisms of Wittgenstein's late philosophy of language and, in particular, of its behaviorist stand. Since the times of Skinner's book and Chomsky's review, the field of language acquisition has expanded dramatically in both empirical coverage and theoretical sophistication. Also in this case, many textbooks are available today that cover the field in its many dimensions. Despite the many progresses, however, the question of what is provided by the environment and what by the learner's cognition in the process of language learning remains a matter of heated debates. Our only claim in the context of this debate is that both aspects are relevant, hence, neither can be truly dismissed in favor of the other. We will reiterate this claim in Chap. 30, when criticizing semantic internalism.

## References

- Austin, J. L. (1962). *How to do things with words*. Oxford: Oxford University Press.
- Canfield, J. V. (1996). The community view. *The Philosophical Review*, 105, 469–488.
- Chomsky, N. (1959). A review of B. F. Skinner's verbal behavior. *Language*, 35(1), 26–58.
- Chomsky, N. (1969). Some empirical assumptions in modern philosophy of language. In E. Nagel, S. Morgenbesser, & P. S. M. G. White (Eds.), *Philosophy, science, and method*. New York: St. Martin's Press.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics*. 3: *Speech acts* (pp. 41–58). New York: Academic Press.
- Kripke, S. A. (1982). *Wittgenstein on rules and private language*. Harvard: Harvard University Press.
- Searle, J. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.
- Skinner, B. F. (1957). *Verbal behavior*. Acton, MA: Copley Publishing Group.
- Strawson, P. F. (1964). Intention and convention in speech acts. *The Philosophical Review*, 73, 439–460.
- Wittgenstein, L., & Anscombe, G. E. M. (1953). *Philosophical investigations*. London: Blackwell.



# Chapter 18

## Meaning and Reality



### Things as They Are and Things as We Talk About Them

We move now to a second objection to semantic externalism. As we saw, semantic externalism is the thesis that natural language meaning is grounded in the external world of material things. If we take the external world to be the natural realm of material objects, it follows from the thesis of semantic externalism that the way we talk about things corresponds, ultimately, to the way things are in the material world. According to some, this is incorrect as there is an irreducible difference between things as they really are and things as we talk about them.

Consider, as a first example, the common noun “book”. What is its meaning? According to the thesis of semantic externalism, the meaning of “book” is a function of what books are in the world of material things and not of what speakers think it is. Consider, then, the following sentences, all comprising the word “book”.

- (1) The book has 540 pages and weights 3 lb.
- (2) The book took a decade to write
- (3) The book is easy to read
- (4) The book sold a million copies

Intuitively, sentence (1) conveys reference to a book regarded as a material object, characterized by physical features such as weight and number of pages. Sentences (2) and (3), however, do not refer to the book as a material object. Rather, they refer to the book as a *content*, what the book says. Similarly, sentence (4) does not inform us about the book as a material object. Rather, it refers to the book as something that is sold in copies, a general abstract *prototype* that comes in a multiplicity of particular material tokens. Now, what sort of natural material object is that which has a certain weight and number of pages, takes some time to write, is more or less easy to read, and is sold in copies? That is, what sort of natural object is that which is, at once, a physical entity, a content, and a prototype?

The most natural reaction to these examples is to say that the noun “book” does not merely refer to books as they are but, rather, as we, the speakers, *conceive* of them, that is, as rich conceptual entities that comprise aspects and dimensions beyond their purely physical qualities. Semantic externalism, however, warns us precisely against this type of explanation. The meaning of a linguistic expression, it submits, does not depend on factors that are internal to the mind of its speakers. Admitting that the meaning of “book” relies, even in some part, on the conceptualization speakers do of the relevant objects would amount to giving up semantic externalism.

A similar example is offered by (5).

- (5) Saint Petersburg is Russia’s second most populous city, an important commercial port on the Baltic Sea, and a magnificent example of neoclassic architecture.

The sentence in (5) offers a description of some characteristics of the city of Saint Petersburg. According to it, the object of reference of the noun “Saint Petersburg” satisfies, at once, the properties of being the second most populous city in Russia, an important commercial port, and a magnificent sample of neoclassic architecture. Clearly, these three properties are true of different aspects of the city. The first concerns its population, the people who live in it. The second is true of the city as a port, a commercial institution and a set of infrastructures. The third applies to the city in terms of its buildings and public spaces. We can easily think of even more properties that apply with equal validity to the city of Saint Petersburg. Amongst the major Russian cities, it is the closest to Europe. It has a vibrant cultural life and is regarded by many as Russia’s cultural capital. It is home to the Hermitage and of many foreign consulates. In all these cases, we refer to the city in yet other ways: as a geographical location, a cultural hub, the host of museums and political institutions. If semantic externalism is correct, we are bound to conclude that, in the external world of natural things, there exists a natural object that is, at once, a group of people, a port, a conglomerate of buildings, a geographical location, a cultural actor, and the host of cultural and political institutions. On the face of it, one could reasonably react that, in the objective world, there is no such thing as the city of Saint Petersburg. There is soil, water, people, buildings, squares, factories, offices, political institutions, laws, customs, and so on. All different elements that together contribute, in all sorts of different ways, to what we call Saint Petersburg.

Another class of expressions that are especially problematic for the thesis of semantic externalism in the context of our current discussion is that of *fictional* names. These are names, such as “Sherlock Holmes” or “John Watson”, whose referents do not even exist in the physical world. Also problematic are those expressions whose objects constantly change their physical constitution. Consider the name “Aleppo”. According to historians, Aleppo is one of the oldest cities in the world. Evidence suggests its first settlement dates back to 6000 BC. Since its original foundation, the city has undergone many changes, including, sadly, those caused by the recent war. Compared to the original city, Aleppo is today an altogether different object. As Theseus ship, each one of its constitutive parts has been changed,

modified, replaced in the course of its long journey through the millennia. Yet, we refer to the original city by the same name—"Aleppo"—that we use to refer to the current city, even though the two objects are, in terms of their strictly material properties, entirely different. In fact, we use one and the same name to refer to each stage the city has gone through, moment after moment, in the course of its millennial life. But how can we say that the name "Aleppo" has maintained its meaning across this millennial life if the material object it is about has not remained the same?

These examples suggest that there are fundamental differences between things *as they are* and things *as we talk about them*. Language appears to presuppose a certain degree of conceptualization of the reality it is about.

## Things that Count and Things that Don't

An important case study in the context of the objection raised in this chapter concerns the distinction between the two classes of common nouns known as "mass" nouns and "count" nouns. This distinction is found in many languages, including English. English mass nouns include names such as "milk", "sunshine", "software", "advice", and "knowledge"; count nouns include names such as "chair", "person", "program", "suggestion", and "belief". The distinction is of value to linguists because the names in the two classes show distinct grammatical behaviors.

Firstly, count nouns can be pluralized, whereas mass nouns resist pluralization. "Chairs", "persons", "programs", "suggestions", and "beliefs" are all grammatical expressions of English. "Milks", "sunshines", "peoples", "softwares", "advices", and "knowledges" are, conversely, all ungrammatical.

Secondly, count nouns can combine with so-called numeral quantifiers, such as "one", "two", "three hundred", and "a million", whereas mass nouns cannot. "One chair", "two persons", "three hundred programs", "four thousand suggestions", and "a million beliefs" are all grammatical expressions of English. "One milk", "two people(s)", "three hundred software(s)", "four thousand advice(s)", and "a million knowledge(s)" are all ungrammatical, irrespectively of whether they are in the singular or plural form.

Thirdly, count nouns can combine with so-called "individuate" quantifiers, such as "each", "every", "few", "several", and "many", whereas mass nouns cannot. "Each chair", "every person", "few programs", "several suggestions", and "many beliefs" are all grammatical expressions of English. On the other hand, "each milk", "every sunshine", "few software(s)", "several advice(s)", and "many knowledge(s)" are all ungrammatical.

Finally, mass nouns can combine with expressions of measure, such as "a glass of", "a bit of", "a lot of", "much", and "two Gigabits worth of", whereas count nouns cannot. "A glass of milk", "a bit of sunshine", "two Gigabits worth of software", "a lot of advice", and "much knowledge" are all grammatical expressions of English, whereas "a bit of chair", "a lot of person", "two gigabits worth of program", and "a lot of suggestion" are not.

What determines the different grammatical behavior of the two classes of nouns? An intuitive explanation is that the difference between the two classes is, ultimately, semantic. The nouns in the two classes display different grammatical behaviors because they refer to different types of objects: Mass nouns refer to *masses* whereas count nouns refer to *countable objects*. Mass nouns refer to stuff, count nouns to things. The terminology that is commonly adopted to designate the two classes—“mass nouns” and “count nouns”—reflects this hypothesis.

Masses and countable objects, as they are typically understood, are distinguished by two properties: *divisiveness* and *cumulativity*. Divisiveness is the property of all those objects whose essential qualities are proper to them as a whole as well as any of their subparts. Divisiveness applies to mass objects but not to countable ones. Imagine you have a certain quantity of milk, say, a glass. If you pour some milk out of the glass, what remains in the glass is still milk. This is because milk is a mass. In contrast, imagine you have a chair and start breaking its pieces apart. At some point you will not have a chair anymore and this is because a chair is a countable object. Cumulativity is the property of all those objects whose essential qualities apply to them as well as to any combination of them. It applies to mass objects but not to countable ones. If you add more milk to the milk that is already in your glass, you still have milk as a result. Conversely, arbitrarily dismantling and recombining any two chairs into a new object gives no guarantee that the result will be again a chair.

Together, divisiveness and cumulativity capture the fundamental difference between masses and countable objects. The grammatical differences between mass nouns and count nouns follow from these properties. Masses do not come in plural tokens. Rather, they cumulate into larger masses of the same type. Hence, mass nouns resist pluralization. Also, masses do not come naturally in units that can be counted. For this reason, mass nouns resist numerals and individuating quantifiers, but can be turned into measurable units by means of expressions of measures, such as “a glass of” or “two spoons of”.

At first, this approach seems on the right track. The distinction between masses and countable object can be regarded as the outcome of a natural distinction between their referents—a distinction, that is, that depends on the material properties of the objects the two classes of names refer to. It is intuitive to think that “water” is mass but “chair” count because water, as a material object, satisfies divisiveness and cumulativity, whereas chairs do not.

Yet, as soon as we broaden our scope to other, less trivial cases, we find that things are not so straightforward. To begin with, when we compare different languages, we easily find nouns that, despite referring to the same object, qualify as mass in some languages but count in others. For example, the Spanish word for hair, “pelo”, is a mass noun, like its English counterpart “hair”. Yet, the corresponding Italian word, “capello”, shows all the grammatical properties of a count noun. It can be pluralized, combined with numerals, and so on. Similarly, whereas the English “strawberry” is a count noun, its Russian counterpart, “klubnika”, is mass. In fact, even within the same language, it is not difficult to find pairs of nouns that have referents that share common material properties but, nonetheless, are classified one

as count and the other as mass. For example, English “spaghetti” qualifies as a mass noun whereas “noodle” is count. Similarly, “garlic” and “rice” qualify as mass nouns whereas “onion” and “bean” qualify as count. Furthermore, it is not difficult to find mass nouns that refer to objects that obviously qualify as countable or count names that refer to objects that are obviously masses. For example, although “furniture” is a mass noun, furniture clearly does not satisfy either divisiveness and cumulativeness. It cannot be infinitely divided or combined and still remain furniture. Infinitely dismantling your furniture or infinitely assembling it with other furniture is likely to take you to a point where you cannot call it furniture anymore. Similar examples are provided in English by the nouns “cutlery”, “clothing”, “equipment”. Finally, the distinction between mass and count applies also to abstract nouns. For example, “belief” and “suggestion” are count nouns whereas “knowledge” and “advice” are mass. Abstract objects, by their very nature, do not have physical properties. Yet, they are also classified as being either mass or count.

These observations show that the objects we talk about are identified as being mass or count quite arbitrarily and, to a large extent, irrespectively of whether they are themselves, in a material sense, mass or count. The lesson we draw is that the grammatical distinction between mass and count nouns that is attested across natural languages cannot be reduced to a corresponding material distinction between their objects of reference. The way speakers talk about things as mass or countable objects is not a function of the way things actually are.

More generally, we have a further argument against semantic externalism. Linguistic meaning cannot be explained solely and exclusively as a function of the external world of material things.

## References and Remarks

The objection to semantic externalism we reviewed in this chapter is due to Chomsky and most explicitly formulated in his book *New Horizons in the Study of Language and Mind* (Chomsky 2000; see also Chomsky 1986; Pietroski 2018).

The distinction between mass and count nouns was first described on the basis of morphological and syntactic criteria by the linguist Otto Jespersen (Jespersen 1924). The analysis of the distinction on the basis of the properties of the notions of divisiveness and cumulativeness began with Godehard Link (Link 1983) and was then further developed by Gennaro Chierchia (Chierchia 1998a, b). Their analysis regards objects as algebraic structures known as *lattices* and derives the notions of divisiveness and cumulativeness as algebraic properties of such structures. The discussion of the distinction between mass and count nouns in the context of the topic of natural language metaphysics is due to Francis Jeffrey Pelletier. The discussion we have presented in the second part of this chapter is based on his essay “Descriptive metaphysics, natural language metaphysics, Sapir-Whorf, and all that stuff: Evidence from the mass-count distinction” (Pelletier 2011).

## References

- Chierchia, G. (1998a). Plurality of mass nouns and the notion of 'semantic parameter'. In S. Rothstein (Ed.), *Events and grammar* (pp. 53–103). Dordrecht: Kluwer.
- Chierchia, G. (1998b). Reference to kinds across languages. *Natural Language Semantics*, 6, 339–405.
- Chomsky, N. (1986). *Knowledge of language*. New York: Praeger.
- Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
- Jespersen, O. (1924). *The philosophy of grammar*. London: Allen & Unwin.
- Link, G. (1983). The logical analysis of plurals and mass terms: A lattice- theoretical approach. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, use and interpretation* (pp. 303–323). Berlin: de Gruyter.
- Pelletier, F. J. (2011). *Descriptive metaphysics, natural language metaphysics, Sapir-Whorf, and all that stuff: Evidence from the mass-count distinction*. In *Baltic International Yearbook of Cognition, Logic and Communication* Vol. 6.
- Pietroski, P. M. (2018). *Conjoining meanings. Semantics without truth-values*. Oxford: Oxford University Press.

## Chapter 19

# Meaning and Ideas



In the previous chapter, we have seen how the way we talk about things cannot be reduced to the way things are in the external world of material things. This observation offers an argument against the thesis of semantic externalism, which, in contrast, argues that the way we talk about things must be traced back to the way things are in the external world of natural things.

We shall now observe that semantic externalism also commits us to the stronger claim that the *essential* features of linguistic meaning are, ultimately, features of the external world of natural objects. In Chap. 14, we distinguished the metaphysics that is practiced by philosophers, which investigates the essential features of the world, from the natural language metaphysics that is practiced by semanticists, which investigates the essential features of linguistic meaning. We should notice now that semantic externalism, if correct, obliterates such distinction by enforcing the identity between what is essential about things as we talk about them and what is essential about things as they are.

In part I, we considered a model of interpretation of natural language that is organized around a few essential ingredients. Of central importance amongst these ingredients are the notions of objects and possible worlds. Is it plausible to trace these notions back to the external world of natural things, as semantic externalism demands? Against this possibility, one may point out that objects and possible worlds are—at least in the form we have characterized them—*abstract* entities and not concrete ones. As such, they cannot be reduced to material features of the external world. Objects, as we have intended them in the context of Montague's theory of compositionality, are abstract *mathematical quantities*. The only feature that qualifies them as such is that they can be members of the mathematical structures we call sets. Possible worlds, as we saw in Chap. 11, are *hypothetical states of the world*, which, by definition, represent states of the world that are alternative to the world as it happens to be. Both these notions are of an explicit abstract nature. Surely, they cannot be pinned down to the bare material properties of the external world.

Before mounting an argument against semantic externalism on the basis of these observations, it is informative to return for a moment to the history that brought Montague to the development of his model-theory and, in particular, to the definition of its essential properties. As we saw in part I, Montague borrowed the essential structure of his model of interpretation of natural language from that originally developed for the interpretation of formal languages, following a practice that began with Frege. It was Frege, indeed, who initiated the use of purely formal and, therefore, abstract notions to provide a formal language that is both precise in its interpretation and general in its application. When doing so, however, Frege was resting on a long-standing philosophical tradition according to which abstraction is a dimension of the real world. This philosophical tradition begins with Plato and maintains that the external world of natural things is not just crude matter but also comprises, in and by itself, an abstract dimension.

## A World of Ideas

To familiarize ourselves with Plato's view, let us look at semantic externalism from a different angle. Until now, we have regarded semantic externalism as a hypothesis on the essential properties of natural language meaning. In the context of formal languages, however, semantic externalism plays a very different role. It is not a hypothesis but a *desideratum*—a requirement that a formal language must satisfy in order to be of scientific value. Suppose we devised a formal language to refer to a model of objects—such as the one we developed back in Chap. 6. Suppose that we want to ensure that the language is scientifically trustworthy, in the sense that it describes the model *objectively*. To achieve this goal, what we need to do is to ensure that the referents of the expressions of the language are decided on the basis of external, objective criteria and the process of interpretation is not influenced by any subjective factors. To be objective, the language must be grounded in reality and represent things as they really are, not as they are interpreted from the subjective standpoint of a thinking subject. In practice, to ensure the objectivity of the formal language, we must enforce semantic externalism on it.

Notably, this reasoning becomes problematic if we try to apply it to the formal language we perhaps trust the most, mathematics. We trust mathematics for two reasons. Because it is *objective*—we take the expression “ $2 + 2 = 4$ ” to express something objectively true, whose value is unaffected by our subjective judgment of it—and because it is *abstract*—we take the mathematical expression “ $2 + 2 = 4$ ” to describe a relation that is valid irrespectively of the concrete objects it applies to—that two and two makes four is true irrespectively of whether it concerns apples or stones. Objectivity, on the one hand, is what makes mathematics reliable. Abstraction, on the other, is what makes it general in scope of application. Objectivity and abstraction, however, are not so easy to reconcile with one another. Objectivity demands that mathematics is about objects—mathematical quantities—whose value is independent from the subjective inclinations of a thinking mind. Abstraction



demands that mathematics is not about concrete objects but abstract ones. But how can something abstract not be the product of a thinking mind?

This problem was already a concern of Plato, who advanced the following answer: Mathematical objects are abstract but, nonetheless, real. Plato believed that the external world is not just a brute mass of unorganized matter. He believed that there is a systematic order within the structure of reality and that this order transcends the material dimension. According to Plato, our universe is populated by two types of objects, concrete and abstract. Concrete objects are material and bound to a location in space and time. Abstract objects are immaterial, immutable, and eternal but, nonetheless, belong to the external world because they exist independently of the minds who think of them. As Plato called these abstract entities *ideas*, his view came to be known as *platonic idealism*.

A more contemporary answer, to the question still in the spirit of Plato's original intuition, would be that we live in a mathematical universe. Like platonic ideas, mathematical objects have no material substance, yet they belong to the fabric of the universe. Reality is not a haphazard amass of raw matter. The material constituents of the physical universe are organized around a common mathematical framework that is abstract but, nonetheless, real because it would exist even if there was no mind to think about it.

The view of mathematical objects as platonic ideas explains how mathematics manages to be, at once, abstract and objective. On the one hand, the reality it refers to is made of immaterial relations that hold independently of the objects they apply to. These are the relations that determine that adding two mathematical quantities to two mathematical quantities delivers a total of four mathematical quantities, irrespectively of what these quantities are quantities of. On the other hand, these immaterial relations are real as they belong to the real world out there and not to the inner realm of the mind.

Plato's idealism, in any of its many incarnations, also offers us a framework to explain the abstract dimension of linguistic meaning in a way that does not endanger semantic externalism. When developing his notion of formal language, Frege was firmly resting on the philosophical tradition set by Plato. In his mind, the essential ingredients of his formal language were reflecting the external world both abstractly and objectively. These are the same essential ingredients that grounded Montague's attempt at providing a model-theory of natural language meaning. If, like Plato and Frege, we accept that the world around us encompasses an abstract dimension, the fact that natural language meaning also encompasses an abstract dimension stops being a threaten to semantic externalism.

## Grasping Ideas

This Platonist view of linguistic meaning, however, also encounters some fundamental problems. Two of them are especially relevant to our current discussion.

The first is that it is not clear that all that is abstract in linguistic meaning is abstract in a platonic sense. Take books. As we saw in the previous chapter, the notion of a book obviously involves some degree of abstraction. A book is not only a material object but also a content and a prototype. Yet, it is not as obvious that this type of abstraction is of the same sort that characterizes mathematical objects.

The second problem is a problem that has interested platonic idealism since its very inception. It is a problem of epistemology: If ideas are immaterial and, yet, external to our minds, how do we, human beings, come to know them? We cannot observe them as we do with material objects, because they are immaterial, and we cannot perceive them through our mind's eye, because they are external objects, not psychological ones. So, how can we even realize their existence? How do they manage to become part of our cognitive lives?

In mathematics, this conundrum is typically solved by pointing out that mathematical objects can be acknowledged through pure reasoning because they hold their properties as a matter of logical *necessity*. The propositions of mathematics are, in effect, all necessary propositions. That two plus two makes four is not simply true, but also necessarily so. It is simply impossible to conceive of a logically coherent state of the world, no matter how different from the actual one, that would make it false. In Chap. 12, we saw that necessary propositions are also a priori—or so many thinkers have contended. If they are a priori, we do not need to investigate their truth by observation. We can prove it by sole reasoning, by demonstrating that they could not be otherwise. In effect, this is what mathematicians do. They construct proofs that a certain mathematical proposition is necessarily true by demonstrating that, if it were false, it would lead to a logical contradiction. Hence, if we are able to acknowledge the existence of mathematical objects despite the fact that they are abstract and immaterial it is because they have necessary properties, which we can identify a priori on the basis of pure reasoning.

Whether this solution is satisfying is a matter of intense debate. From our perspective, it is important to observe that, even if it were correct, it would not apply as easily to natural language. The things we talk about in natural language, no matter if concrete or abstract, are, by and large, of a contingent nature, and not immutable eternal objects with necessary properties. In fact, the vast majority of the propositions we express in natural language are a posteriori contingencies, not a priori necessities. They describe what happens to be the case, rather than what is necessarily so and could not be otherwise. But then, if the propositions of natural language are not logical necessities and yet abstract, how do we manage to grasp them? Speakers do enjoy a rich array of cognitive relations with linguistic meaning. Propositions are not only what people express and understand through language. They are also, as we saw back in Chap. 13, the objects of propositional attitudes—what people believe, think, know, hope, expect, want, remember, and so on. If the propositions we express in natural language are abstract objects of the platonic sort, but also are, by and large, contingent and not necessary, then we are left with no explanation of how they can play any role in the cognitive lives of the thinking subjects who entertain them.

It is certainly important to realize that linguistic meaning encompasses a significant degree of abstraction. Yet, the abstraction we express in natural language cannot be all explained away in the terms of some form of Platonic idealism.

## References and Remarks

The most known formulation of Plato's view is his allegory of the cave, which is presented in the VII book of the *Republic*. Many scholars trace the roots of Plato's view to the mathematical and geometric discoveries of Pythagoras. Pythagoras was the first to observe that the spatial and quantitative properties of reality transcend its physical constituents and are immaterial, immutable, and eternal. This finding had a powerful, in fact shocking impact on Pythagoras and his colleagues. It was precisely because of the aura of mysticism that surrounded the discovery of an abstract dimension to reality that their philosophy was perceived more akin to a religion and their circle to a sect.

Frege's anti-psychologist position is most clearly expressed in his *Die Grundlagen der Arithmetik* (Frege 1884) and then in his *Grundgesetze der Arithmetik* (Frege 1893/1903). The same position is defended by Russell in his article "Meinong's Theory of Complexes and Assumptions" (Russell 1904). More recently, Frege's and Russell's platonic view of propositions has been criticized by Scott Soames in his essay "Why the traditional conceptions of propositions can't be correct" (chap. 3 of King, Soames, & Speaks 2014) and his book *Rethinking Language, Mind, and Meaning* (Soames 2015).

## References

- Frege, G. (1884). *Die Grundlagen der Arithmetik. Eine logisch mathematische Untersuchung über den Begriff der Zahl*. Breslau: Verlag von Wilhelm Koebner.
- Frege, G. (1893/1903). *Grundgesetze der Arithmetik*. Jena: Verlag Hermann Pohle, Band I/II.
- King, J. C., Soames, S., & Speaks, J. (2014). *New thinking about propositions*. Oxford: Oxford University Press.
- Russell, B. (1904). Meinong's theory of complexes and assumptions. *Mind*, 13(1), 204–219. 336–354, 509–524.
- Soames, S. (2015). *Rethinking language, mind and meaning*. Princeton: Princeton University Press.

## Chapter 20

# Meaning and Subjectivity



### Private Language (And Why People Use It Anyways)

The third and last objection we will consider against semantic externalism will occupy us for the next few chapters. The objection is that, on the face of Wittgenstein's argument, private language appears to be perfectly possible. Describing the external world is without doubt an important function of language. Many would agree, nonetheless, that people also use language to describe what goes on in their minds, their inner sensations, emotions, memories, feelings, and so on.

As we know, Wittgenstein's argument against private language is meant to demonstrate that a logically private language is impossible. Yet, when one is confronted with Wittgenstein's example and asked to imagine taking note of a private sensation on a diary, one rarely reacts by saying, "Hey, wait a minute! What you are asking me is logically impossible!" It is in fact not rare at all for people to take similar private records, for example, for therapeutic reasons.

Natural languages themselves are quite well endowed when it comes to expressing private sensations. Instead of introducing a new symbol for each new sensation, as Wittgenstein suggests, an English speaker could simply write, in plain English, "my foot tickles" or "I feel hot". Expressions such as these are abundant in all natural languages and they seem to be doing precisely what Wittgenstein's argument prohibits: referring to the privacy of speakers' minds. By Wittgenstein's argument, such expressions should be simply impossible. Yet, we like to think they are not.

## Ouch!

Wittgenstein himself was well aware of this type of objection. He believed, however, that it is based on a mistaken view of meaning. In his mind, the expressions that we intuitively qualify as referring to private objects, such as sensations, feelings, emotions, and so forth, do not really refer to private objects. Rather, we should regard them as external *manifestations* of the inner psychological lives of speakers. As we saw, according to the theory of language as use, language must be understood as one of the actions humans perform in interaction with their environment. It is in such interaction, and only there, that we should observe it. In the context of the theory of language as use, expressions of private psychological states should be understood in the same way. Individuals learn them to *signal* their internal states to the environment. As children cry to signal that they are in pain, so adults learn to say “I am in pain” to achieve the same goal.

There is a whole set of problems related to this view and a vast literature on the topic, expounding a range of sometimes very different positions. One important problem, to which we will return, is that expressions of internal subjective states are not always in the mouth of those who experience them. A sentence such “I am in pain” is easy to account for in Wittgenstein’s terms. Speakers, as we saw, use it to signal their inner state of pain to their environment. But what about the sentence “Frida is in pain”, as uttered, for example, by Kazimir? This sentence expresses the fact that somebody other than the person speaking is in pain. But how is this possible if pain is a private sensation? How can somebody signal an inner psychological state on somebody else’s behalf?

A further problem is, of course, that Wittgenstein’s solution works only within the more general framework of the theory of meaning as use, whose shortcomings we discussed already in Chap. 17.

## Subjectivity

In the coming chapters, we will address this issue by observing, from a more detailed linguistic perspective, how the notion of *subjectivity* unfolds in natural language meaning. This investigation will take us to a forced choice between two mutually exclusive options: either accepting the Wittgensteinian view that all there is to natural language is behavior or accepting that the model of interpretation of natural language contains an irreducible element of subjectivity.

Before embarking in this enterprise, it is useful to clarify our understanding of the term “subjectivity”. In particular, it is useful to distinguish between two different senses of the term. Subjectivity can characterize a type of knowledge, in which case we refer to it as *epistemic* subjectivity, and it can characterize a type of entity, in which case we refer to it as *ontological* subjectivity.

Epistemic subjectivity has to do with the way something is acknowledged. We talk about objective knowledge when the knowledge does not depend on a particular individual's perspective. We have objective knowledge of how many pages there are in this book, how tall Mount Everest is, and when Kazimir Malevich was born. In contrast, we talk about subjective knowledge when the knowledge depends on the knower's perspective. That the book is enjoyable, for example, is subjective. It depends on the individual entertaining such knowledge.

Ontological subjectivity has to do with the nature of things. In philosophy, ontology is the study of things as they are. We say that things such as rocks, chairs, and trees are ontologically objective, because they belong to the external world. If, as we discussed in the previous chapter, we agree with Plato, we can also regard mathematical entities as being objective. Conversely, things such as sensations, emotions, memories, perceptual impressions, and so on, belong to the realm of the mind and are, therefore, ontologically subjective.

In the coming chapters, we will see that, if we wish to capture natural language meaning in terms of reference to a model, we must accept that this model contains elements of both epistemic and ontological subjectivity. According to the theory of meaning we have developed so far, the function of language is to describe a model of objects. We shall see that this idea must be enriched in two respects. Firstly, language does not simply describe a model. It always presents it from a *perspective*. In this respect, it offers an epistemically subjective description of the model. Secondly, the model described by natural language comprises, amongst its essential elements, also entities that are subjective in the ontological sense.

By and large, our inquiry will focus on one tiny, yet powerful word: "I".

## References and Remarks

Wittgenstein's account of expressions of sensation as behavioral avowals is presented in §244 of his *Investigations* (Wittgenstein & Anscombe 1953):

How do words refer to sensations? – There doesn't seem to be any problem here; don't we talk about sensations every day, and name them? But how is the connection between the name and the thing named set up? This question is the same as: How does a human being learn the meaning of names of sensations? For example, of the word "pain". Here is one possibility: words are connected with the primitive, natural, expressions of sensation and used in their place. A child has hurt himself and he cries; then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behaviour. "So you are saying that the word 'pain' really means crying?" On the contrary: the verbal expression of pain replaces crying, it does not describe it.

The philosophical literature on this issue is truly vast. A good starting point is Stewart Candlish and George Wrisley's entry "Private Language", in *The Stanford Encyclopedia of Philosophy* (Candlish & Wrisley 2014). A significant difficulty in extending Wittgenstein's behavioral account to expressions of other individuals' sensations is Wittgenstein's own stance about the inaccessibility of other sentient

individuals' minds. This point is made in §302 of the Investigations (Wittgenstein & Anscombe 1953):

If one has to imagine someone else's pain on the model of one's own, this is none too easy a thing to do: for I have to imagine pain which I don't feel on the model of pain which I do feel. That is, what I have to do is not simply to make a transition in the imagination from pain in one place to pain in another. As from pain in the hand to pain in the arm. For it is not as if I had to imagine that I feel pain in some part of his body. (Which would also be possible.)

Pain-behaviour can indicate a painful place but the person who is suffering is the person who manifests pain.

The argument suggested in this passage maintains that the experience of a sensation is an essential property of the sensation itself, hence it is impossible to conceive of a sensation as it is experienced by somebody else because that would mean missing an essential feature of the sensation. In this case as well, the literature offers a number of diverging interpretations of Wittgenstein's original remarks. Central contributions to this debate are Norman Malcolm's 1954 review of *Philosophical Investigations* (Malcolm 1954) and his "Knowledge of other minds" (Malcolm 1958) and Kripke's postscript to his 1982 essay—entitled "Wittgenstein on other minds" (Kripke 1982). In part III of this book, we will consider a view, grounded on the scientific study of perception, whereby conscious experience is an arbitrary sign of the human organism's inner states (sensorial, emotional, and so forth), whose significance for the organism's central cognitive systems (such as memory and reasoning) has been established in the course of human evolution for the sake of enhancing the organism's chances of survival in its environment. According to this view, the meaningfulness of conscious experience does not depend on its being experienced but on its phylogenetic foundations. This view, as we will see, makes conscious experience not only a possible object of linguistic reference but also something that can be attributed to other individuals on the basis of the fact that it is interpretable within the framework provided by a shared evolutionary history.

The distinction between the epistemic and ontological senses of the notion of subjectivity that we discussed in this chapter is from Searle's book *Mind, Language, and Society* (Searle 1998, pp. 44–45).

## References

- Candlish, S., & Wisley, G. (2014). Private language. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Stanford, CA: Stanford University. <https://plato.stanford.edu/archives/fall2014/entries/private-language/>.
- Kripke, S. A. (1982). *Wittgenstein on rules and private language*. Harvard: Harvard University Press.
- Malcolm, N. (1958). Knowledge of other minds. *The Journal of Philosophy*, 45, 969–978.
- Malcolm, N. (1954). Wittgenstein's philosophical investigations. *The Philosophical Review*, 63, 530–559.
- Searle, J. (1998). *Mind, language, and society*. New York: Basic Books.
- Wittgenstein, L., & Anscombe, G. E. M. (1953). *Philosophical investigations*. London: Blackwell.

# Chapter 21

## Meaning and Indexicality



### Indexicals

The world is full of objects and each of us is one of them. Most natural languages rely on specific expressions to allow speakers to refer to themselves. These are *first-person* pronouns, such as the English pronoun “I”. Our challenge for this and the coming chapters will be that of characterizing the meaning of these expressions.

Informally, a first-person pronoun, such as “I”, refers to the person who pronounces it or writes it. If Kazimir says “I”, it means Kazimir. If Frida says “I”, it means Frida. The referent of “I” is, therefore, *variable*. This feature alone represents already a significant challenge for the way we have conceived meaning so far. Up to this point, we have assumed that the reference of a simple expression is always stipulated in the lexicon and is, therefore, *fixed*. By that we mean that it is not affected by the environment in which it is used. The reference of the proper name “Kazimir Malevich”, to pick an example, is the individual Kazimir Malevich. This referent is stipulated once and for all in the lexicon of the language and it is not affected by who utters the name, whom it is uttered to, when it is uttered, or where it is uttered. This, however, is not the case with “I”. The first-person pronoun belongs to a class of expressions whose reference changes systematically depending on the context in which they are used. The reference of “I”, in particular, depends on who is talking (or writing, depending on the mode of communication). When uttered by Kazimir, it refers to Kazimir. When uttered by Frida, it refers to Frida.

Other English expressions that also vary systematically with their context of use are “you”, “here”, “now”, “today”, “yesterday”, “tomorrow”, to mention only a few. The reference of “you” depends on who is the addressee of the talking. The reference of “here” depends on where the talking is taking place. The reference of “now” depends on when the talking is taking place. And so forth. In all these cases, the reference of the expression is not established by means of a lexical stipulation. It is rather determined by some coordinate or, as it is sometimes called, *index* of the



context in which the expression is used. This is why these expressions are often referred to as *indexicals*.

Indexicals are of special interest for the debate between semantic externalism and semantic internalism because of their ambivalent nature. On the one hand, indexicals appear to be inextricably bound to their context of use, which provide the necessary conditions for their reference. On the other hand, indexicals appear to contribute an epistemic value that goes beyond their mere reference. As we will conclude from the discussion in this and the following chapters, a satisfactory analysis of the meaning of indexicals will require us to identify a common ground between those aspects of meaning that are bound to the external world of natural things and those that belong to the inner realm of the mind.

## Indexicality and Reference

A good way to start analysing the meaning of indexicals is by considering the challenges they pose to the theory of meaning as reference—the thesis, that is, that the meaning of a linguistic expression is to be identified with its reference.

A first challenge has to do with the knowledge that enables speakers to use them and understand them proficiently. As we have just seen, simple expressions are normally assigned a referent in the lexicon and this referent is all speakers need to know in order to use them and understand them appropriately. To learn the meaning of “Kazimir Malevich” we must learn what its reference is. Once we have this knowledge, we can use and understand the expression. This is not the case with indexicals, however. Indexicals do not come with a reference, assigned to them in advance, they receive a reference in the moment in which they are used. “I”, for example, does not refer to anything until the moment somebody utters it. Then, once uttered, we know its referent is the person uttering it.

So, if it is not its referent, what is it that speakers know when they know how to use and understand an indexical? What is it, for example, that speakers of English know when they know how to use and understand “I”? To answer these questions, we must identify something that is common to all uses of “I”, which speakers can learn and which enables them to use it and understand it whenever the occasion arises. This, however, cannot be its reference because that changes all the time.

A second challenge posed by indexicals to the view that meaning is reference is that indexicals appear to contribute more meaning than their sole reference. Compare (1) to (2).

- (1) I am a painter
- (2) Kazimir Malevich is a painter

Consider a scenario where (1) is uttered by Kazimir Malevich himself. In such a scenario, “I” refers to Kazimir Malevich himself. Now, if the meaning of “I” is

equivalent to its reference, we must also conclude that (1) expresses the same meaning as (2)—the proposition that the individual Kazimir Malevich belongs to the set of painters.

This seems wrong. Sentence (1) contributes more information than just the proposition that Kazimir Malevich is a painter. This is confirmed by the fact that there are plausible scenarios where Kazimir Malevich may agree with (2) while, at the same time, disagreeing with (1).

To exemplify this possibility, we will use the assistance of a very peculiar character, amnesiac Kazimir. The real Kazimir Malevich died in 1935 at the age of 57. The last years of his life had been bitter, as the Stalinist regime had established a harsh policy against any form of abstract art. Malevich's work was regarded as "bourgeois" and derided by Russian critics. Many of his paintings were confiscated and Malevich was banned from producing and exhibiting abstract art. Let us imagine, for a moment, that things had gone differently. Kazimir Malevich is still alive. He has lived a long and prosperous life and his work is appreciated by art critics all around the world. Because of his very old age, however, he now suffers from amnesia. More often than not, he has no recollection of his past, his work, or his talent for painting. Sometimes, he does not even remember his own name.

Consider, now, the following scenario. One day, amnesiac Malevich enters an art gallery, while on a state of profound amnesia. He spends hours admiring the beautiful abstract paintings on display and carefully reads all the information available about their author, a Russian painter who, he learns, responds to the name of Kazimir Malevich. In this scenario, the two sentences in (1) and (2) do not appear to be completely equivalent. Whereas Kazimir would agree that (2) is true, he would still not agree that (1), as uttered by him himself, is also true. As far as amnesiac Malevich can tell, it is indeed the case that the individual known under the name of Kazimir Malevich is a painter and, therefore, (2) is true. However, in the same scenario, he would not agree that his utterance of (1) is true for the obvious reason that, even though he acknowledges that Kazimir Malevich is a painter, he still fails to realize that Kazimir Malevich is he himself.

This scenario replicates a scenario proposed in a number of varieties in the philosophical literature to demonstrate that sentences such as (1) and (2) have different meanings. Although both sentences express the information that the individual Kazimir Malevich is a painter, (1) conveys the further information that the individual uttering the sentence acknowledges the individual his sentence is about as himself.

Importantly, these observations extend to other indexical terms. We can, for example, construct a similar scenario with the second-person pronoun "you". Suppose you are now the person visiting the gallery. You look at the paintings and learn all sorts of things about Kazimir Malevich. Then, you express your knowledge of the painter to a visitor standing nearby. Unbeknownst to you, the visitor you are talking to is Kazimir Malevich himself. In this context, you may agree to utter (2), but you would not agree to utter (3).

## (3) You are a painter

Having knowledge that Kazimir Malevich is a painter is not enough to authorize you to utter (3). To do so, you must hold also knowledge of the fact that Kazimir Malevich is the person you are talking to. “You”, similarly to “I”, does not only convey its reference. It also expresses the fact that the speaker acknowledges such reference in a certain way.

The conclusion is that there is more to the meaning of an indexical than its mere reference. If the meaning of an indexical was simply its reference, we would expect (1), (2), and (3) to be equivalent ways of expressing the same information. On the contrary, the scenarios we have considered show us that these sentences are not equivalent.

## Indexicals as Descriptions

As a first attempt at addressing the challenges posed by indexicals, we will discuss the view that indexicals are not referential expressions but, in fact, definite descriptions in disguise. In Chap. 10, we discussed the difference between two types of nominal expressions, proper names and definite descriptions. There, we saw that “Kazimir” and “the person with a hat” perform a similar function in that they both allow speakers to single out an individual entity. In fact, they may even be used to single out the same individual entity. For example, in a contest where Kazimir is the person with the hat, the two nominal phrases can both be used to talk about Kazimir. We also saw, however, that the two noun phrases perform their function in different ways. Whereas proper names *refer* to individual entities, descriptions *describe* them. The meaning of a proper name is its referent, which is stipulated lexically. The meaning of a definite description is a description—a property that has the ability to uniquely identify an individual in the circumstances at hand.

It is also useful recalling that the difference between proper names and descriptions can be captured by a possible world semantics. As we saw in Chap. 11, the referent of a proper name remains constant across different possible states of the world, whereas the referent of a definite description is free to change from possible world to possible world, depending on the entity that happens to satisfy the relevant description. On a parallel with the distinction between proper names and definite descriptions, we have also distinguished between two types of propositions: *de re* propositions, which are about individual entities, and *de dicto* propositions, which are about descriptions.

In the previous section, we have considered the limitations of identifying the meaning of “I” and other indexicals with their reference. Perhaps, these limitations are the result of incorrectly assuming that “I” and the other indexicals express reference to a *res*—like proper names do—and, as a consequence, contribute *de re* propositions. What if, instead, we treated them as a definite description contributing to *de dicto* propositions? After all, definite descriptions are strikingly similar to

indexicals in that the individual entity they single out varies with the circumstances, depending on which entity happens to satisfy their description. As “the person with a hat” picks up the individual that uniquely satisfies the property of wearing a hat in the circumstances at hand, so “I” picks up the individual that uniquely satisfies the property of being the person currently speaking in the circumstances at hand.

These observations suggest that indexicals should be regarded as equivalent to definite descriptions and, accordingly, as contributing *de dicto* propositions. We can regard “I” as a shorthand for the definite description “the speaker” or, at any rate, any description of the form “the *P*”, where *P* is a property that uniquely identifies the individual that is currently producing the pronoun, in speaking, writing, or any other relevant mode of communication. We can similarly analyse “you” as equivalent to the definite description “the addressee”, “now” as “the time of speaking”, “here” as “the place of speaking”, and so forth.

Such an approach provides us with a solution to the challenges presented above. The fact that “I”, like the other indexicals, picks up different individuals in different circumstances is now a natural consequence of the fact that “I” is equivalent to a definite description. When the person that uniquely satisfies the description “the speaker” is Kazimir, “I” picks up Kazimir. But when the person that uniquely satisfies the property of being the speaker is Frida, “I” picks up Frida. The fact that “I” picks up different individuals depending on who is talking is, after all, not so different from the fact that “the person with a hat” picks up different individuals depending on who is wearing the hat.

The descriptive approach also explains what it is that speakers know when they know how to use and comprehend an indexical. If “I” is equivalent to a definite description, then to know its meaning is not to know a referent, but, rather, an identifying description. This description is the element of the meaning of “I” that is constant across its different uses and what speakers must learn in order to be able to use it and interpret it correctly.

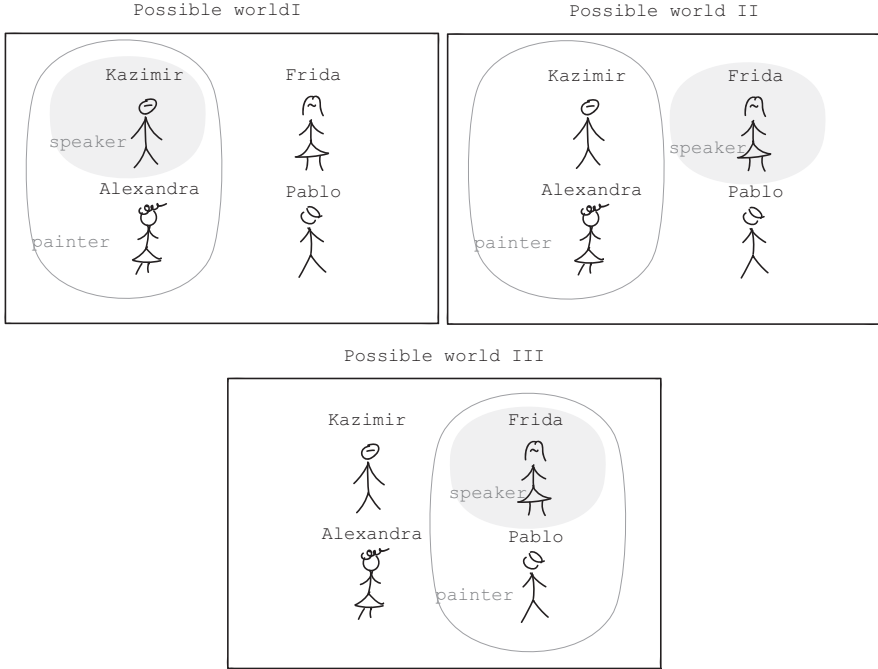
The descriptive approach to indexicals also explains why sentences (1) and (2), repeated below, are not equivalent even though they are both uttered by Kazimir.

- (1) I am a painter
- (2) Kazimir is a painter

Since “I” is now equivalent to the description “the speaker”, (1) is now equivalent to (4), not (2).

- (4) The speaker is a painter

Sentences (2) and (4) express different propositions. Sentence (2) expresses a *de re* proposition, whereas (4) expresses a *de dicto* proposition. More precisely, (2) refers to the set of possible worlds where Kazimir is a painter, whereas sentence (4) refers to the set of possible worlds where the individual who happens to be the speaker is a painter. These two sets of possible worlds overlap partially but not exactly, as there are possible worlds where one sentence is true but the other is false. To appreciate this point, consider the three possible worlds in Fig. 21.1. Possible world I is a possible world where Kazimir is the speaker and he is also a painter.



**Fig. 21.1** A collection of three possible world. Possible world I is a possible world where Kazimir is the speaker and he is also a painter. Possible world II is a possible world where Frida is the speaker, but Kazimir is a painter, whereas Frida is not. Possible world III is a possible world where Frida is the speaker, not Kazimir, and she is also a painter, whereas Kazimir is not a painter

Possible world II is a possible world where Frida is the speaker but Kazimir is a painter, whereas Frida is not. Possible world III is a possible world where Frida is the speaker, not Kazimir, and she is also a painter, whereas Kazimir is not a painter. To tell whether a possible world belongs to the proposition expressed by (2), we must first identify the individual Kazimir and then check if he is a painter in that world. When we do this with the possible worlds in the figure, we find that worlds I and II belong to the proposition expressed by (2), but world III does not. In worlds I and II, Kazimir does belong to the set of painters, whereas in III he does not. Conversely, to tell whether a possible world belongs to the propositions expressed by (4), we must first find the individual that is the speaker in that world and then check if that individual is a painter in that world. When we apply this to the possible worlds above, we find that worlds I and III belong to the proposition expressed by (4) but II does not. In worlds I and III the individual that is the speaker is also a painter. In the case of I, Kazimir is the only individual who is both the speaker and a painter. In the case of III, that individual is Frida. In the case of II, conversely, the individual that is the speaker is not a painter.

The two sentences are, therefore, verified by different sets of possible states of the world. Sentence (2) is true in all the possible worlds where Kazimir belongs to

the set of painters, regardless of whether he is also the speaker. Sentence (4) is true in all those possible worlds where the individual that uniquely satisfies the property of being the speaker belongs to the set of painters, regardless of whether he is Kazimir.

## I am not Here Now

The descriptive approach to the meaning of indexicals has significant advantages. Yet, it also suffers from fundamental problems.

The first argument we will consider against the descriptive theory of indexicals is due to the philosopher and logician David Kaplan. It applies to “I” as well as to the other indexicals. Kaplan’s observation is that indexicals do not contribute de dicto propositions at all. When we say “I am a painter”, our intent is not to predicate a relation between the property of being the speaker and that of being a painter. Rather, we genuinely mean to talk about a *res* and, accordingly, to contribute a de re proposition.

To prove his point, Kaplan makes ingenuous use of some simple yet puzzling sentences, such as (5).

(5) I am speaking

Here is what makes sentence (5) an interesting puzzle. In Chap. 10, we saw that the definite determiner “the” expresses a relation between two properties. According to this analysis, a sentence of the form *the + noun + predicate*—such as “the painter sings”—is true if, and only if, the following two conditions are met: The property denoted by the noun—in our example, “painter”—is a property that uniquely identifies a single object; and such property (description) entails the property denoted by the verb—in our example, “sings”. According to this analysis, the sentence “the painter sings” is true if, and only if, there is only one painter in the relevant model and that individual is also a member of the set of those who sings. In the framework developed in part I, we characterized “the” as expressing the function in (6).

(6)  $X \rightarrow Y \rightarrow 1$  if  $X$  contains exactly one entity and  $X$  is a subset of  $Y$ , 0 otherwise

The function in (6) is a relation between two sets  $X$  and  $Y$ , which is true whenever  $X$  contains exactly one entity and  $X$  is a subset of  $Y$ . When applied to the sentence “the painter sings” it delivers the truth if, and only if, the set of those who paint contains exactly one element and is a subset of the set of those who sing.

The relation expressed in (6) enjoys a specific mathematical property, *reflexivity*. We say that it is reflexive because, if we feed it two identical arguments, it is bound to deliver the truth. That is, whenever we pick two identical sets as our  $X$  and  $Y$ , the function automatically delivers the truth-value 1. In practice, reflexivity means that sentences such as “the painter is a painter” or “the book is a book”, where noun and predicate refer to the same sets of objects, all express logically necessary and a priori propositions. They cannot be false and their truth can be decided on the basis

of pure reasoning, by simply observing that they share the common logical format “the  $P$  is  $P$ ”.

This observation provides us with a powerful tool for testing the descriptive theory of indexicals. Let us assume that “I” is indeed a short-hand for a definite description. Let us further assume, for the sake of the demonstration, that this description is “the speaker” (but notice that the argument we are about to formulate can be constructed with any description we chose to be appropriate when characterizing the meaning of “I”). If we are correct in these assumptions, we predict that (5) expresses a proposition that is both necessary and a priori. If, in fact, “I” is identical to “the speaker”, then, we must conclude that (5) is identical to (6).

(6) The speaker is speaking

Now (6), is a sentence of the form “the  $P$  is  $P$ ”, which, according to the reasoning above, is both necessary and a priori. Hence, if (5) is identical to (6), (5) must also be necessarily and knowable a priori.

Is this prediction correct? Do speakers of English attribute to (5) a meaning that is both logically necessary and a priori? This seems obviously incorrect. Clearly, (5) does not correspond to a logical necessity. The fact that the person who turns out to be speaking is speaking is not a matter of logical necessity. Indeed, it is not difficult to conceive of the possibility that someone saying “I am speaking” could have been doing something else than speaking. That is, the property of being speaking is not a necessary property of an individual uttering the sentence “I am speaking”. If we agree with this observation, then, we have a straightforward argument against the view of indexicals as descriptions.

Before rejecting the descriptive analysis, we should observe that there is, nonetheless, something rather peculiar about (5). In effect, it is difficult, if not impossible, to conceive of a scenario where one utters the sentence without making it automatically true (try it!). This effect appears to be a consequence of the fact that uttering the sentence is a sufficient condition for making it true. In other words, even though sentence (5) does not express a necessary proposition, it nonetheless appears to share some of the properties of a priori statements. A priori statements, as we saw, are statements whose truth can be known in advance, on the basis of pure reasoning and without empirical investigation of the facts they describe. Sentence (5) can be regarded as expressing an a priori proposition, although in a somewhat more practical sense: We cannot know its truth in advance but we can tell in advance that it will express a true proposition *whenever uttered*.

All in all, the descriptive theory of “I” makes a correct prediction as well as an incorrect one. Clearly, the theory ends up describing a sentence such as (5) as expressing a logical necessity and that is just not the way speakers of English understand it. This demonstrates that speakers do not mean the sentence to express a relation between properties. They genuinely interpret it as the contingent attribution of a property to an individual. At the same time, we must also grant to the descriptive approach to indexicals that it correctly predicts that the sentence is passible of an a priori judgement, in the sense that speakers of English can tell in advance that the

sentence is true whenever uttered, even without knowledge of the exact circumstances in which it is uttered.

Another famous example used by Kaplan to make the same point is sentence (7).

(7) I am here now

Sentence (7) is similar to sentence (5) in that it is impossible to utter it without making it automatically true. Nonetheless, the sentence does not express a logical necessity. It is perfectly reasonable to conceive of a possible alternative to the actual world where the person uttering (7) is not at the place and time of utterance of the sentence. This is demonstrated by the fact that its negation is not only a perfectly legitimate sentence of English, but also one that surged to enormous success starting from the 1970s, when answering machines became a common piece of domestic technology:

(8) I am not here now

The fact that (8) can indeed be true, shows that (7) can indeed be false and, therefore, that it is not a logical necessity.

All in all, Kaplan's puzzling sentences show us that indexicals behave as genuinely referring expressions, contributing contingent *de re* propositions. This is problematic for the descriptive approach to indexicals, which, instead, treats indexicals as descriptions and the propositions they contribute to as *de dicto*. The same sentences, however, enjoy a peculiar form of practical *a priori*. They are not true of any state of the world, hence not logically necessary, but they can, nonetheless, be predicted to be true whenever uttered.

## Find Yourself

There is a second argument against the descriptive approach to indexicality, which is more deeply intertwined with the specific first-personal nature of the propositions that one entertains about oneself when using the first-person pronoun. To refer to these propositions, the philosopher David Lewis introduced the term *de se*, which literally means "about oneself". It is to him and the philosopher John Perry that we owe the argument we are about to illustrate.

The goal of this argument is to question the very idea that a description can allow one to identify one as oneself amongst the objects of the world. According to the descriptive theory, "I" is equivalent to a definite description of the form "the *P*" where *P* is a property that univocally identifies the person uttering the pronoun. It is the property that univocally identifies Kazimir when Kazimir is the person speaking and Frida when Frida is the person speaking. What is this property exactly? Above, we have simply assumed that a description such as "the speaker" would do the job and that, if it does not, there must be out there some other property *P* that would eventually do the job. According to Lewis and Perry, however, there is no such property. That is, there is no property that can ultimately do the job of allowing one to



univocally individuate one as oneself. Hence, there is no description that allows one to say “since that object satisfies that description, then that object must be me”. There is something about identifying oneself as such which is of a more primitive logical nature than the mechanism of acknowledging a property in an object.

One may immediately react to this by claiming that there are of course descriptions that uniquely identify one as oneself—for example, “the person that is me”, “the person that is identical to me”, or “the person that I know for a fact to be me”. These descriptions, however, offer no real solution to the problem at hand because they themselves contain an occurrence of a first-person indexical—“I” or “me”. As our goal is to identify a description that captures the meaning of “I”—or, for that matter, “me”—we cannot simply resort to a description that, in itself, contains the indexicals “I” or “me”. We must identify a more primitive description, which in itself does not contain the notion we are trying to explain.

To exemplify the argument of Lewis and Perry, we shall make use again of amnesiac Malevich. Our fictional character is still in a profound state of amnesia. He is very impressed by his visit at the museum and the paintings by this Russian painter called Kazimir Malevich. He decides, henceforth, to visit a library where he reads as much information about Kazimir Malevich as he can. He reads about his early years, his education, his paintings, his family, his relationships, his thoughts, his theories, his later years, even his amnesia. Will amnesiac Kazimir ever realize that the painter he is reading about is, in fact, he himself? Indeed, it is possible that, at some point, amnesiac Malevich encounters a piece of information that awakens his memory and brings back the awareness that Kazimir Malevich is, in fact, he himself. However, it is also perfectly possible that he continues indefinitely to acquire information about him himself without ever realizing the information is in fact about him himself. Even assuming the library is incredibly vast and contains all information there is about Kazimir Malevich, it is still possible that none of this information makes amnesiac Kazimir conclude, as a matter of logical deduction, that it must be about him himself. Even if—as it would happen in a post-modern novel—Kazimir were to read a perfectly detailed description of himself in his current state and circumstances, it would still be possible that this is not sufficient information for Kazimir to conclude with certainty that the description is about him himself.

Scenarios such as this one suggest the following conclusion. There is no property that, once recognized in an object, allows an individual to conclude, as a matter of *logical necessity*, that the object is the individual herself. There is no property that Kazimir may learn about himself that will grant him, as a matter of strict logical deduction, the conclusion that that property cannot be about no one but him himself. Kazimir may learn about his age, place of birth, history, ideas, and many other things. None of this information is such that Kazimir cannot but conclude that “since the object I am learning about has this property, then this object must be me!” Whatever it is that makes us recognize ourselves as such—whatever it is that makes us say “that’s me!”—it is not the realization that a certain object satisfies a certain property. It is something significantly more primitive than that. The logical appara-

tus that allows us to acknowledge properties in objects is secondary to the more primitive capacity of acknowledging ourselves as such.

This peculiarity of the first-person was already noticed by Frege. He observed that people are always presented to themselves in a “special and primitive way”, in which they are presented to no one else. With Frege, we will refer to this “special and primitive way” in which one is presented to oneself as *self-acquaintance*. The arguments by Perry and Lewis, which we exemplified in this section, demonstrate Frege’s intuition that self-acquaintance is “special and primitive” by showing us that the notion of self-acquaintance cannot be reduced to the notion of a property.

Of course, these observations spell trouble for the descriptive account of “I”. As there is no descriptive property that allows one to identify oneself as such, there is also no definite description that can safely replace “I”. Above, we introduced the descriptive analysis of “I” as a way to capture the fact that its reference varies with the circumstances of its use as well as the fact that its meaning cannot be reduced to its bare reference—when Kazimir says “I” we do not only understand that he is referring to a certain individual but also that he is identifying that individual as him himself. The argument we reviewed in this section shows us that this process of identification cannot be reduced to the process of identifying the object that satisfies a description. The special relation of self-acquaintance that holds between Kazimir, as the individual speaking, and Kazimir, as the individual referred to, when Kazimir says “I” can just not be captured by a description.

Notice that the argument extends to all other indexicals. All indexicals, in fact, are ultimately dependent on the first-person and, therefore, self-acquaintance. “You” is the individual “I” am talking to. “Here” is the place where “I” find myself when talking. “Now” is the time at which “I” find myself when talking. And so on. Indexicals are all essentially rooted in this primitive relation of self-acquaintance.

The challenge in addressing the meaning of indexicals is, therefore, that of reconciling their referential value with their cognitive significance. Both the referential and the descriptive analysis are, in this respect, insufficient. The referential analysis focuses on the reference of indexicals, but disregards their mode of identification. The descriptive analysis, focuses on the mode of identification, but disregards their reference and, ultimately, fails to capture the special and primitive nature of the relation of self-acquaintance. The next chapters will be dedicated to providing a framework that captures all these different dimensions of indexicals.

## References and Remarks

The modern semiotic notion of indexical is due to the logician and philosopher Charles Sanders Peirce. We refer the interested reader to Albert Atkin’s essay “Peirce on the index and indexical reference” (Atkin 2005).

The picture of indexicality we have offered in this chapter is simplified in several respects. In particular, we have presented expressions such as “I”, “you”, “here”, and “now” as *pure* indexicals—that is, expressions whose meaning is determined as an

index of their circumstances of utterance. This is an oversimplification with a number of exceptions. “I”, for example, can refer to an individual other than the speaker when used in a direct report, as in “Frida said, ‘I am a painter’”. In this case, “I” does not refer to the person uttering the sentence but to Frida. For an in-depth discussion of the cognitive and communicative dimensions of the relation between the phenomena of direct and indirect reports, we refer the interested reader to Alessandro Capone’s monograph “The pragmatics of indirect reports” (Capone 2016). “You” is possible of a generic use, as in the sentence “if you want to become a good painter, you have to do a lot of practice”. In this sentence, “you” can be understood not as referring to the addressee but to a generic individual. “Here” is possible of a demonstrative interpretation. Imagine, for example, Frida pointing at a city on a map and saying “I want to go here”. In this case, “here” can be understood as referring to a place other than the place of utterance. Finally, “now” can refer to a time other than the time of utterance when used in the so-called historical present tense—as, for example, in the sentence “it is now 1915 and Kazimir Malevich publishes his artistic manifesto”. The discussion in this chapter overlooks these ambiguities and, instead, focuses on the purely indexical use of these expressions. Our presentation also overlooks the class of expressions known as ‘demonstratives’. These are simple expressions such as “this” and “that” or complex expressions such as “this painting” or “that man”. Like indexicals, demonstratives receive their semantic value from their context of use, although they are also typically accompanied by a gesture—sometimes called a demonstration, hence the name demonstratives.

Kaplan’s observations on indexicality (as well as his logical framework for capturing their meaning, to which we will return in the next chapter) were first formulated in his articles “On the Logic of Demonstratives,” (Kaplan 1978), “Demonstratives”, and “Afterthoughts” (Kaplan 1989a, b). His observations as well as the theoretical framework he developed apply to both indexicals and demonstratives.

The observation that some indexical sentences—such as “I am speaking” or “I am here now”—are made true by the very fact that they are uttered relates them to a class of statements known as *performatives*. Performatives were brought to the attention of linguists and philosophers by Austin in his *How to Do Things With Words* (Austin 1962). They are statements such as “I pronounce you husband and wife” or “I hereby declare that the information given is true and correct to the best of my knowledge and belief”. As Austin famously observed, the value of these statements cannot reside in their truth-conditions, because those would be trivially verified by the very act of pronouncing the sentences. The simple fact of pronouncing the sentence “I pronounce you husband and wife” is a sufficient condition for making the sentence true. Their value resides, rather, in the performance of the action of uttering them. On the face of the similarities, we should observe that Austin’s observations do not readily apply to the cases involving indexicals. Whereas it is true that the sentences “I am speaking” and “I am here now” are verified by the very act of pronouncing them, it is not as true that they contribute trivial truth-conditions.

The argument we presented in the section “Find yourself” is based on work by Perry—in particular, his essay “The problem of the essential indexical” (Perry 1979)—and Lewis—in particular, his essay “Attitudes de dicto and de se” (Lewis 1979). The term de se was introduced by Lewis in the same article. Frege’s observation that individuals are presented to themselves in a “special and primitive way” is formulated in his essay “Der Gedanke” (Frege 1918-1919).

## References

- Atkin, A. (2005). Peirce on the index and indexical reference. *Transactions of the Charles S. Peirce Society*, 41(1), 161–188.
- Austin, J. L. (1962). *How to do things with words*. Oxford: Oxford University Press.
- Capone, A. (2016). *The pragmatics of indirect reports: Sociophilosophical considerations*. Heidelberg: Springer.
- Frege, G. (1918-1919). Der Gedanke. Eine logische Untersuchung. *Beiträge zur Philosophie des deutschen Idealismus*, 1, 58–77.
- Kaplan, D. (1978). On the logic of demonstratives. *Journal of Philosophical Logic*, 8, 81–98.
- Kaplan, D. (1989a). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). Oxford: Oxford University Press.
- Kaplan, D. (1989b). Afterthoughts. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 565–614). Oxford: Oxford University Press.
- Lewis, D. (1979). Attitudes de dicto and de se. *Philosophical Review*, 88, 513–543.
- Perry, J. (1979). The problem of the essential indexical. *Noûs*, 13, 3–21.

# Chapter 22

## Meaning and Context



### Perspective

Let us imagine Kazimir Malevich beginning his speech at an event organized in St. Petersburg by a local union on March, 22nd 1913 by uttering (1).

(1) I am here today to talk to you about the future of art

The scenario that (1) describes is, factually speaking, perfectly equivalent to that described by (2).

(2) Kazimir Malevich is in St. Petersburg on March, 22nd 1913 to talk to the audience at the Union of Youth about the future of art

So why not uttering (2) instead of (1)? What is special about (1)? The idea we will explore in this chapter is that (1) does not only contribute a reference, but also a *perspective* – the perspective of a certain speaker, Kazimir Malevich, towards a certain addressee, the audience at the Union of Youth, in a certain place, St. Petersburg, and at a certain time, March 22nd 1913.

What do we mean by perspective? It is useful to draw a parallel with the figurative arts. The notion of perspective originates with the Renaissance. During that time, perspective was predominantly understood as a *mathematical* notion—a system of principles of projective geometry that allow translating a three-dimensional space into a two-dimensional one. Objects exist in a three-dimensional space. They have height, width, and depth. A painting, conversely, is two-dimensional. It has only height and width. Perspective is the set of mathematical principles that translate the three dimensions of the original into the two dimensions of the painting that depicts it. In time, however, it became clear that there is another dimension to perspective, one that has to do with the *psychological* principles that allow an observer to interpret a two-dimensional image as the depiction of a three-dimensional object.

The mathematical and psychological dimensions of visual perspective are both important. The first focuses on the properties of the object that is *observed*, which

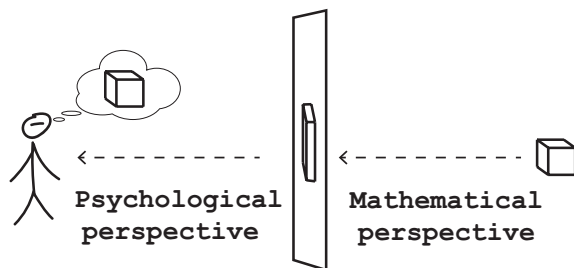
causes the illusion of a third dimension. The second focuses on the *observer*, the perceiver of the illusion. The two dimensions are exemplified in Fig. 22.1.

For a time, these two ways of understanding visual perspective were regarded as distinct, if not contradictory. This changed in the 1960s, when the art historian Ernst Gombrich demonstrated that the two views are not in opposition, but, in fact, describe two complementary aspects of a bi-dimensional phenomenon. According to Gombrich, “the painted third dimension is the co-operative creation of artist and spectator.” (Adamson 2016, p. 6). Perspective emerges from the interaction between the geometric properties of the painting and the perceptual skills of the observer. Both elements, observer and observed, are necessary ingredients of a comprehensive account of what visual perspective is and how it works.

On a parallel with Gombrich, our strategy for capturing perspective in language and, in this way, provide an account of the meaning of indexicals, will be twofolded. Our first step, which we will undertake in this chapter, will consist in anchoring the model to a *context*—the center of the perspective—and, in this way, provide a mathematical model of the relation between the meaning of indexical expressions and the context in which they are used. The resulting framework, as we will appreciate in the next chapter, extends to the notion of meaning well beyond the domain of indexicals. The second step, which we will undertake in Chap. 24, will be to understand the cognitive relation that holds between language users and the meaning of the expressions they use. This will require us to investigate the primitive and special notion Frege identified as self-acquaintance.

## Context

The goal of a model, at least in the way we have understood it so far, is that of offering a description of a state of affairs. The model we have developed in part I offers a description that organizes a set of individual entities around properties and relations of different sorts. We will now enrich our model by adding a further element—the



**Fig. 22.1** The two dimensions of visual perspective: the mathematical dimension to the right—the set of geometric principles that translate a three-dimensional object into a two-dimensional image—and the psychological dimension to the left—the set of cognitive principles that allow the observer to interpret a two-dimensional image as the depiction of a three-dimensional object

*vantage point* from which the model is constructed. We will refer to this new element in the model as the *context*.

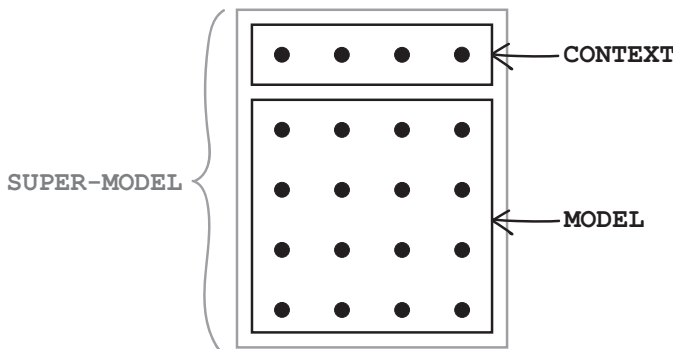
The notion of context, in the sense that is relevant to our discussion, was introduced in model-theoretic semantics by Kaplan in his seminal article “On Demonstratives”. In our model, we can think of the context as a set of objects including the individual speaker, the point in time at which the speaker is speaking, the point in space at which the speaker is speaking, and the individual (or individuals, in case she is addressing more than one person) to whom the speaker is speaking. Simply put, a context is a package of coordinates that tells the interpreter who the speaker is, when and where she is speaking, and to whom she is speaking. More formally, we can treat a context as an *ordered* set of indexes:

(3)  $\langle s, t, p, a \rangle$

The notation in (3) represents an ordered set of four values. Value  $s$  is a symbol for the individual that is speaking,  $t$  for the time of utterance,  $p$  for the place of utterance, and  $a$  for the addressee (or group of addressees, in case the speaker is addressing more than one person). In a nutshell, the context of a model of interpretation is a set of indexes representing the circumstances in which the interpretation is taking place.

It is important to understand that, formally speaking, the context is not an element *in* the model. Rather, context and model, as they are constructed in Kaplan’s theory, contribute together, as two distinct elements, to the logical structure that allows speakers to interpret language. We can think of the combination of context and model as a *super-model* which comprises both the model as we have conceived it so far—with its objects and properties—and the context—with its speaker, addressees, time of utterance, and place of utterance. This is exemplified in Fig. 22.2.

As we can think of the model as a *photograph* of the world, so we can think of the context as a description of the *photographer*, the individual taking the picture



**Fig. 22.2** The model of interpretation of indexicals proposed by David Kaplan. It encompasses both a model—intended, as we have done so far, as a set of objects organized around their properties and relations—and a context—the coordinates of the speaker, addressee, place of utterance, time of utterance, etc

together with her spatio-temporal coordinates. The super-model is the sum of these two elements: photographer and photograph, observer and observed.

## Kaplan's Logic of Indexicals

Kaplan's framework relies on the notion of context as a formal element in the logical structure for the interpretation of language. The goal is to model the mathematical relation that holds between the context in which an expression of the language is used and the reference it takes in the model. So far, we have maintained that the meaning of a linguistic expression is the object it refers to in the model. Accordingly, we have constructed our theory of meaning as a system that pairs expressions  $E$  in the language with corresponding objects  $O$  in the model, as in Fig. 22.3. We can symbolize the function performed by our system of interpretation as in (4)—that is, as a function that takes an expression  $E$  of the language as its input and delivers an object  $O$  of the model as its output.

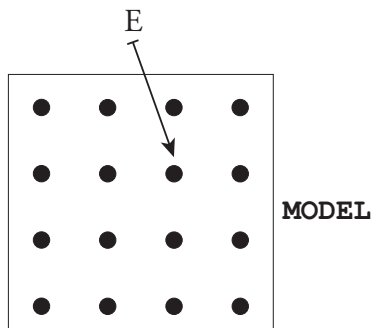
$$(4) E \rightarrow O$$

The discussion from the preceding chapter has taught us that the function in (4) is problematic for indexicals, because, as we saw, indexicals refer to different objects in different circumstances. Despite being simple expressions, their reference is not stipulated in the lexicon of the language but varies with their context of use. So, for them, we are unable to express a *constant* function that determines their reference.

Kaplan's notion of context allows us to address this problem. In the system of interpretation he devised, indexicals receive their reference in the model through the mediation of the context. As exemplified in Fig. 22.4, an indexical expression  $I$  receives its reference by means of a two-step function. First, it is anchored to the context. Then, it is assigned a reference. We can exemplify the function depicted in the figure as in (5).

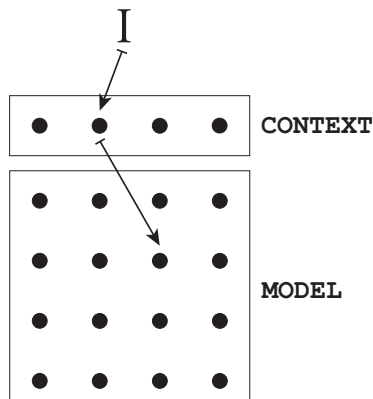
$$(5) I \rightarrow (C \rightarrow O)$$

**Fig. 22.3** An expression  $E$  is assigned an object in the model as its reference





**Fig. 22.4** An indexical  $I$  is assigned a reference in the model through the mediation of the context



According to (5), an indexical expression  $I$  is first interpreted as a function ( $C \rightarrow O$ ) that takes a context  $C$  as its input and delivers an object  $O$  in the model as its output. It is only when a context  $C$  is applied as the input to this function, that the function delivers an object  $O$  as the reference of the indexical in the model. According to this analysis, indexicals do not refer directly to objects in the model. They rather do so through the mediation of a function which, when applied to a context, provides a reference in the model.

To familiarize ourselves with this framework, let us apply it to some actual indexical expressions. Let us begin by observing that different indexicals provide different indexical functions as they individuate different objects in their context. We can define these different functions by relying on the formulation of context we have provided above. As we saw, a context of utterance is an ordered set of values  $\langle s, t, p, a \rangle$ , which are, respectively, the individual  $s$  uttering the expression, the time  $t$  of utterance, the place  $p$  of utterance, and the individual or individuals  $a$  to whom the expression is addressed. On the basis of this, we can interpret the indexical “ $I$ ” as the function that maps a context  $\langle s, t, p, a \rangle$  into the object in the model that is the speaker  $s$  in the context. This function is formulated explicitly in (6).

$$(6) \quad C \rightarrow s$$

Consider, as an example, a context where Kazimir is talking to Frida in St. Petersburg, on March 22nd. This is represented formally as the ordered set in (7).

$$(7) \quad \langle \text{Kazimir, St. Petersburg, March 22}^{\text{nd}}, \text{Frida} \rangle$$

If we apply this context to the function in (6), we obtain, as its outcome, the value Kazimir:

$$(8) \quad \langle \text{Kazimir, St. Petersburg, March 22}^{\text{nd}}, \text{Frida} \rangle \rightarrow \text{Kazimir}$$

In the context at hand, the speaker in the context is Kazimir. Hence, Kazimir is the object in the model chosen by the function as the reference of “ $I$ ”.

We can easily extend this framework to other indexicals by interpreting them on the basis of similar functions selecting different values from the context. “Here” expresses the function that maps the context into the time of utterance  $t$ , “now” the function that maps the context into the place of utterance  $p$ , “you” the function that maps the context into the addressee  $a$ :

$$(9) \quad C \rightarrow t$$

$$(10) \quad C \rightarrow p$$

$$(11) \quad C \rightarrow a$$

If we apply the context represented in (7) to these functions, we obtain, as their respective outputs, St. Petersburg, March 22nd, and Frida. These are the referents assigned to “here”, “now”, and “you” in such a context.

## Content and Character

Kaplan’s framework reconciles two important aspects of indexicals—the fact that they refer to individual entities, not descriptions, and the fact that their reference varies with the context in which they are used. As desired, the functions deliver different results when applied to different contexts. In a context where Frida is talking to Kazimir, “I” is mapped into Frida and “you” into Kazimir. Conversely, in a context where Kazimir is talking to Frida, “I” is mapped into Kazimir and “you” into Frida. As the reference of indexicals is now a function of their context of use, any relevant variation in their context finds a corresponding variation in their reference. At the same time, the ultimate output of these functions is an object in the model. As desired, indexicals convey reference to a *res*, not a description, and, accordingly, contribute de re propositions. The sentence “I am speaking”, when uttered by Kazimir, attributes the property of being speaking to the individual Kazimir, and whether Kazimir is speaking is a contingent fact of the world, not a logical necessity.

We should also notice, however, that the introduction of the notion of context in our system contributes a new dimension to the notion of meaning. To appreciate this point, consider again the function in (5).

$$(5) \quad I \rightarrow (C \rightarrow O)$$

According to (5), an indexical expression  $I$  receives its reference by means of a two-step procedure. First, it is translated into a function ( $C \rightarrow O$ ) from contexts to objects. Then, when applied to a context, this function delivers an object  $O$  in the model that we can finally regard as the reference of the indexical. In a sense, henceforth, an indexical has two meanings. One is the relation it expresses between the context and the model. The other is the object it refers to in the model. In his writings, Kaplan refers to the first notion as the *character* of the indexical and to the second as its *content*. The character is the function that maps the context of use of the indexical to its reference in the model. The content is the referent of the indexical in the model. Which one of these two notions should we regard as the

genuine *meaning* of an indexical? The answer is that content and character are both indispensable ingredients of the meaning of indexicals as they contribute different dimensions to it. It is for this reason that Kaplan's theory is often referred to as a *two-dimensional semantics*.

The character of an indexical is important because it informs the interpreter about the relation that links the context in which the indexical is used to the object it refers to. The character of "I", for example, tells us that the referent of the indexical is the speaker in the context of utterance. As such, the character of indexical is the dimension of its meaning that remains constant across its different uses. "I" may pick different objects when used in different contexts, but it always expresses the same character—the function that maps a context  $\langle s, t, p, a \rangle$  into its speaker  $s$ . The character of indexicals is, therefore, what enables speakers to use and interpret indexicals across their different uses. The notion of character also enables us to explain the peculiar form of a priori that we find with a sentence such as "I am speaking". As we saw in the previous chapter, this sentence is made true by the very fact that it is uttered. In Kaplan's logic, this follows straightforwardly from the fact that the character of "I" entails the property denoted by the predicate "to be speaking". No matter what is the context and who happens to be speaking in it, the indexical function will always provide an individual that, by the very fact that it is selected by such character, satisfies the property of being speaking.

The content of an indexical is also important, because it determines the proposition that the indexical contributes. The sentence "I am a painter", when uttered by Kazimir, expresses the proposition that the individual Kazimir belongs to the set of painters. As desired, this is a *de re* proposition that attributes a property to Kazimir regardless of the fact that he is also the speaker in the current context of utterance. This also explains why the sentence "I am speaking" does not express a logical necessity. When uttered by Kazimir, the sentence expresses the *de re* proposition that Kazimir is speaking, which is contingent, not necessary.

Kaplan's theory provides us with the logical structure necessary to reconcile two central aspects of the meaning of indexicals—the fact that their meaning is a function of their context of use and the fact that they contribute *de re* propositions. These two aspects are captured by grounding the interpretation of indexicals on the notions of model and context and by constructing their logic around the two dimensions of content and character.

## References and Remarks

Our brief presentation of the history of perspective in art history is based on Sylvia Adamson's essay "Deixis and the renaissance art of self construction" (Adamson 2016). Gombrich's ideas on pictorial perspective are presented in his books *Art and Illusion* (1960) and *The Image and the Eye* (1982).

Kaplan's logic of indexicals is formulated in his article "On the logic of demonstratives" (Kaplan 1978; see also Kaplan 1989a, b).

## References

- Adamson, S. (2016). Deixis and the renaissance art of self construction. *Sederi*, 16, 5–29.
- Gombrich, E. H. J. (1960). *Art and illusion: A study in the psychology of pictorial representation*. London: Phaidon.
- Gombrich, E. H. J. (1982). *The image and the eye: Further studies in the psychology of pictorial representation*. Oxford: Phaidon.
- Kaplan, D. (1978). On the logic of demonstratives. *Journal of Philosophical Logic*, VIII, 81–98.
- Kaplan, D. (1989a). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). Oxford: Oxford University Press.
- Kaplan, D. (1989b). Afterthoughts. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 565–614). Oxford: Oxford University Press.

## Chapter 23

# Meaning and Causality



In the previous chapter, we have familiarized ourselves with Kaplan's logic of indexicals and, in particular, with the idea that the reference of indexicals is fixed through the mediation of a context. We shall now observe that the function that relates indexicals to their reference is characterized by an element of *causality*.

In effect, according to the framework we have presented in the previous chapter, the interpretive function of indexicals is that of mapping a feature of the material environment in which they occur into their reference. The material environment of use of an indexical becomes a determining factor of its reference. In a sense, the function performed by an indexical is not so different from that performed by a thermometer, a photo camera, or a weight scale. The function of a thermometer is to translate the temperature of its immediate surroundings into a measure of temperature. In this case, the measure provided is a direct effect of the environment on the thermometer. Similarly, the function of a photo camera is to translate the distribution of light in its environment into an image. Also in this case, the image provided by the camera is a direct effect of the material conditions of the environment on the camera. Finally, the weight scale translates the pressure that is put on it into a measure of weight. Again, this result is a product of a causal effect enacted by the environment on the scale. In Kaplan's theory, indexicals are like thermometers. Their function is to detect a feature of the environment in which they are used and translate it into their reference. "I" detects who is the individual speaking, "you" who is the individual spoken to, "here" the point in space where the speaking is occurring, "now" the point in time.

The finding that the meaning of indexicals is related causally to their environment of use deserves special attention in the context of our discussion of semantic externalism. This finding offers us evidence that the meaning of indexicals is inextricably bound to the material conditions of their environment of employment. As we will see in the following chapters, this conclusion is only partially warranted. Our current theory of indexicality is, in fact, still partial. A full account of the meaning of indexicals will demand also an explanation of the "special and primitive"

relation of self-acquaintance indexicals entail and, as we shall see, self-acquaintance has also important consequences for semantic externalism. For the time being, however, we shall observe that the notion of causality plays an important role in linguistic meaning, well beyond the class of indexicals.

## Two Stars, One Planet

A class of expressions where we find that the notion of causality plays a role is that of *proper names*. Kaplan's logic, in fact, turns out to be a useful framework for explaining a classical puzzle in the semantics of proper names, which is originally due to Frege.

Consider the proper names "Hesperus" and "Phosphorus". "Hesperus" is the name used by the ancient Greeks to refer to "the evening star"—the first visible star in the evening. "Phosphorus" is the name used by the same ancient Greeks to refer to "the morning star"—the last star to fade from the sky in the morning. When the two names were introduced, it was believed that the two stars were different celestial bodies. It was later discovered by Babylonian astronomers that the two stars are in fact the same celestial object—the planet Venus. The content of this discovery can be expressed by sentence (1), which, informally, asserts the identity between the celestial body designated by "Hesperus" and the one designated by "Phosphorus".

### (1) Hesperus is Phosphorus

At a closer look, we find that defining the meaning of (1) is a rather puzzling enterprise. So far, we have maintained that the meaning of a proper name is the object it refers to. Let us apply this idea rigorously to the names "Hesperus" and "Phosphorus" and, thus, to sentence (1). If the meaning of "Hesperus" is the object it refers to, then its meaning is nothing but the celestial body we now identify as the planet Venus. Similarly, if the meaning of "Phosphorus" is the object it refers to, then its meaning is also the planet Venus. But then, if this is correct, we must conclude that (1) expresses the proposition that planet Venus is identical to planet Venus and, therefore, that (1) expresses the identity between Venus and itself.

Needless to say, this cannot be right. We all know that (1) is not meant to express the trivial information that an object is identical to itself. The proposition that an object is identical to itself is a *necessary* proposition, no matter what that object is. There simply is no logically coherent possible world where an object is different from itself. By the same token, there is no possible world where Venus is different from itself. This conclusion is clearly problematic. Our intuitions tell us that (1) does not express a logical necessity. Furthermore, if the sentence expressed a logical necessity, we would expect that its truth would be knowable a priori. This seems also incorrect. To prove the truth of this statement, the Babylonian astronomers relied on sophisticated empirical observations rather than pure reasoning.

The puzzle posed by (1) is, henceforth, that if we take the meaning of the two names to be the object they refer to, we incorrectly predict that (1) expresses the necessary a priori proposition that Venus is identical to itself.

## Names as Descriptions

As in the case of indexicals, Frege's puzzle may tempt us to abandon the idea that proper names refer to objects and contribute *de re* propositions and, instead, adopt the view that names are a shorthand for definite *descriptions* and contribute *de dicto* propositions. This view was contemplated by Bertrand Russell, who proposed that proper names are not *directly referential*—that is, they do not refer directly to entities without the mediation of a description—but, rather, definite descriptions in disguise. According to his view, the name “Hesperus” does not refer directly to the planet Venus but is rather a shorthand for the definite description “the first visible star in the evening”. Similarly, “Phosphorus” does not refer directly to the planet Venus but is a shorthand for the definite description “the last star to disappear from the sky in the morning”. Sentence (1) should then be regarded as equivalent to (2).

- (2) The first star visible in the evening is the last star to disappear from the sky in the morning.

This is a welcome result. The proposition expressed by (2) is neither necessary nor a priori. This is because (2) expresses the identity between two different descriptions rather than between two identical objects. As other definite descriptions, “the first visible star in the evening” and “the last star to disappear from the sky in the morning” are free to pick up different referents in different possible worlds, depending on which objects happen to satisfy the descriptions “first visible star in the evening” and “last star to disappear from the sky in the morning”. The proposition expressed by (2), therefore, distinguishes those possible worlds where the two descriptions are satisfied by the same object from those where they are not. This proposition is, henceforth, neither necessary nor a priori. The fact that, in the actual circumstances, the two descriptions pick the same object—the planet Venus—is an accident of the world, not a logical necessity.

Clearly, the view of proper names as descriptions parallels the view of indexicals as descriptions. It is not surprising, then, to find that it also suffers from similar problems. Soon after Kaplan, Saul Kripke raised significant objections to it, which are parallel to those formulated by Kaplan against the view of indexicals as descriptions. Consider again the proper name “Hesperus” and assume, in accordance with the descriptive theory of Russell, that it is equivalent to the definite description “the first visible star in the evening”. If correct, the analysis entails that sentence (3) expresses a logical necessity.

## (3) Hesperus is the first visible star in the evening

It is easy to see why. If the noun “Hesperus” is equivalent to the definite description “the first visible star in the evening”, then (3) must be equivalent to (4).

## (4) The first visible star in the evening is the first visible star in the evening

Sentence (4), however, expresses a logical necessity of the form “the  $P$  is  $P$ ”. Sentence (3) cannot thus be equivalent to (4), since (3), contrary to (4), does not express a logical necessity. What (3) rather tells us is that a certain celestial object—in this case, Venus—has the property of being the first star visible in the evening. This is a *contingent* property of that object and not a property that object holds as a matter of logical necessity. That Venus happens to be the first star visible in the evening is a contingent fact of our universe, not an eternal and immutable property of any logically coherent universe. The problem we run into is henceforth that, if we take “Hesperus” to be equivalent to the definite description “the first visible star in the evening”, we are forced to conclude that whatever it is what we call “Hesperus”, it has the property of being the first visible star in the evening as a matter of logical necessity. Clearly, this is a mistake.

The lesson we learn from Kripke’s observations is that proper names are used by speakers to refer to individual entities directly and independently of their properties and, therefore, they contribute *de re* propositions. They do not express definite descriptions and they do not contribute *de dicto* propositions. They are, to use Kripke’s terminology, *rigid designators*—expressions whose reference remains constant across different possible worlds. When we utter a sentence like (3), what we really mean to say is that a certain *res* has a certain property—that is, that the object referred to by the name “Hesperus” has the property of being the first visible star in the evening. We do not mean to express a trivial logical necessity, as the theory that “Hesperus” corresponds to a description would force us to conclude.

## The Character of Names

The solution to Frege’s puzzle requires, once again, distinguishing the two dimensions of content and character in the meaning of proper names, as Kaplan did for indexicals. The content of a proper name is the object it picks up in its model of interpretation. Its character is the function that grounds its content in its context. What are then the conditions of the context of use of a proper name that determine its content? To answer this question, we must return to a fundamental property of the meaning of proper names.

As we observed as early as in Chap. 5, simple expressions—that is, those expressions that are not the result of the combination of simpler ones—have the meaning they have as a matter of *stipulation*. This, of course, is also true of proper names. There is no necessary reason why the names “Kazimir Malevich” or “Hesperus”



refer to what they refer to. The reason why these names refer to, respectively, the founder of Suprematism and planet Venus is that, at some point in the past, speakers agreed that they would do so. Speakers may agree on the meaning of different proper names for different reasons. The reason why we use “Kazimir Malevich” to refer to the Russian painter, for example, should be traced back to the moment when Kazimir Malevich’s parents decided upon his name. In this case, we can reconstruct a moment in the past when the association between the name and its referent was first established—a *baptism*, during which the Malevich’s decided that the baby they were holding in their hands would be called “Kazimir”. In the case of “Hesperus” it is practically impossible to reconstruct such moment, but we know that it must have happened at some point or another when a community of speakers agreed, perhaps implicitly, that the name would be used to refer to the object they identified in their physical environment as the star they first saw in the evening sky.

Despite these differences, we find that the two names have the meaning they have as the result of the historical chain of causation that traces back their current interpretation to the moment—no matter how far and forgotten—when their object of reference was first identified. The character of a proper name is, henceforth, its history—the causal chain of historical events that determines the name’s current interpretation. As in the case of indexicals, the character of proper names describes the connection that maps their environment of use into their reference. In the case of proper names, however, the relevant contextual features are not to be found in the physical environment where the name is uttered but, rather, in its past—the historical chain of causation that determines its current reference.

By regarding proper names as having both a content and a character, we can address Frege’s puzzle. As names, such as “Hesperus” and “Phosphorus”, have two dimensions to their meaning, so do the propositions they contribute to. Sentence (1), therefore, expresses both a content and character. Its content corresponds to the proposition that the reference of “Hesperus” is identical to the reference of “Phosphorus”. This is, indeed, a necessary proposition, stating the identity between Venus and itself. The character of (1), however, corresponds to the claim that the character of “Hesperus” is equivalent to the character of “Phosphorus”. This is equivalent to the claim that the conditions that determined the content of the name “Hesperus” brought about the same result as the conditions that determined the content of “Phosphorus”. This is a contingent fact, because the conditions that determined that “Hesperus” would be used to refer to planet Venus are different from those that determined that “Phosphorus” would be used for the same purpose. It is therefore a contingent fact—not a necessary one—that the characters of the two names happen to deliver the same content. We see, finally, that sentence (1) is informative after all and not a priori. The informative part of the sentence, however, is not what it says about the content of the two names. Rather, it is what it says about their characters. The sentence does not inform us that the contents of the two names are equivalent. It rather informs us that the characters of the two names are equivalent.

## Character Generalized

The lesson we draw from Frege's puzzle is that proper names are not so different from indexicals as their content is also a function of their context of use. Hence, also proper names have a character. This character corresponds to the causal chain of events that traces the name back to the moment when its reference was originally established.

Proper names, however, are not the only expressions whose content is stipulated. As we saw, it is a general property of all simple expressions that their reference is stipulated. Furthermore, since by the principle of compositionality the reference of complex expressions is a function of the reference of the simple expressions that constitute them, we must conclude that also their reference is, after all, stipulative. If this is correct, we must conclude that, ultimately, all expressions have a character—that is, all expressions receive their content as the product of a causal chain that grounds them into their context of use. Indexicality is, henceforth, a pervasive property of linguistic meaning: The content of linguistic expressions is inextricably bound to the environment in which they occur.

## Necessary A Posteriori and Contingent A Priori

Before concluding this chapter, it is worth pointing out that the observations made by Kaplan and Kripke against the descriptive view of indexicals and proper names come with important philosophical implications. In Chap. 12, we saw that the distinctions between necessary and contingent, on the one hand, and a priori and a posteriori, on the other, have been traditionally regarded as two sides of the same coin. If a proposition is necessarily true, then it is also knowable a priori. If it is a priori, then it is also necessarily true. Similarly, if a proposition is contingent, then it is knowable only a posteriori. If it is a posteriori, then it is also contingent. The strict correlation between these two oppositions is implicit in many philosophical arguments and much formal reasoning. An example is provided by mathematics. As we saw in Chap. 19, mathematical truths can be investigated objectively despite being abstract because they are necessarily true and, therefore, they can be proven on the basis of sole reasoning, that is, a priori.

The case of natural language statements such as (1)—repeated below—and (5) are relevant to this discussion because they demonstrate that the notions of necessary and contingent are in fact distinct from those of a priori and a posteriori.

- (1) Hesperus is Phosphorus
- (5) I am the speaker

Sentence (1), as we saw, expresses a necessary content, corresponding to the proposition that Venus is identical to itself. Yet, it is a posteriori. This is because the fact that the names “Hesperus” and “Phosphorus” refer to the same object is a matter

of contingency, not necessity. This sentence offers, therefore, an example of a necessary statement that is, nevertheless, knowable only a posteriori. Sentence (5), which we discussed in Chap. 21, exhibits the opposite pattern. The sentence expresses a contingent proposition, attributing to the individual who has been contingently identified as the speaker the contingent property of being the person who is speaking. Yet, the sentence appears to behave as a a priori statement: We can tell in advance and without engaging in any observation that the sentence will be true whenever uttered.

Examples of this sort have been used to demonstrate that the distinction between necessary and contingent must be kept apart from that between a priori and a posteriori. One distinction has to do with the facts of the world that make a statement true or false. The other, with how speakers acknowledge the truth or falsity of a statement. These are just two different businesses.

In the context of these observations, the notions of content and character have been put to serve the more ambitious philosophical task of accounting for the distinction between necessary and contingent, on the one hand, and a priori and a posteriori, on the other: Whereas the content of a statement determines whether it is necessary or contingent, it is its character that decides whether it is a priori or a posteriori. We should emphasize that not everyone agrees that this is a proper use of the two-dimensional framework originally developed by Kaplan, whose main goal was, after all, that of expressing the indexical function that relates the content of an expression to its context of use, not that of accounting for the divide between the notions of necessity and a priori.

## References and Remarks

Frege's discussion of identity statements such as "Hesperus is Phosphorus" is from his "Über Sinn und Bedeutung" (Frege 1892). In the article, Frege relies on the discussion of these statements and other considerations to argue for a distinction between two dimensions of meaning to whom he refers as *Sinn* and *Bedeutung*. *Sinn*, commonly translated in English as "sense", is what determines the reference of a name whereas *Bedeutung*, commonly translated as "reference", is, in fact, the reference of the object. Since Frege's original formulation, the notion of *Sinn* has been characterized in different ways in different formal frameworks and has been put to different uses, beyond those originally devised by Frege. We will return to the distinction between *Sinn* and *Bedeutung* in Chap. 30.

The most prominent example of a descriptive theory of proper names is Russell's in his 1905 article "On denoting" (Russell 1905). Various incarnations of the descriptive approach have been defended by a number of linguists and philosophers until recent times. Notable examples are Bach (1981), Burge (1973), Kneale (1962), Searle (1958), Strawson (1959). Kripke's criticism of the descriptive analysis of proper names is presented in his essay *Naming and Necessity* (Kripke 1980). In it, Kripke presents three arguments against the descriptive approach known,

respectively, as the modal, epistemic, and semantic arguments. In this chapter we offered a simplified version of the epistemic argument.

The causal theory of reference (also referred to as the historical theory of reference) is commonly traced back to remarks made by Kripke himself in his 1980 essay (Kripke 1980, p. 96). Yet, the possibility of a theoretical account of the meaning of proper names based on the notion of causation had already been contemplated before Kripke. A notable example is Gareth Evans, who already discusses such a theory in his 1972 article “A Causal Theory of Names” (Evans 1973). In the same article, Evans also presents some critical objections to the theory.

An analysis of the meaning of proper names based on Kaplan’s notion of character is presented by Ulrike Haas-Spohn in her book *Versteckte Indexikalität und subjektive Bedeutung* (Haas-Spohn 1995). Interestingly, Kaplan had already expressed his opposition for treating proper names on the basis of his notion of character. According to him, “those who suggest that proper names are merely one species of indexical depreciate the power and the mystery of the causal chain theory” (Kaplan 1989, p. 562).

The use of the notions of content and character to address the issue of the relation between a priori and a posteriori, on the one hand, and necessary and contingent, on the other, begins with Kaplan himself. In section XVII, corollary 3 of his “Demonstratives”, Kaplan formulates this possibility as follows: “The bearers of logical truth and of contingency are different entities. It is the character (or, the sentence, if you prefer) that is logically true, producing a true content in every context. But it is the content (the proposition, if you will) that is contingent or necessary” (Kaplan 1989, p. 539). Many philosophers have adopted two-dimensional models to address issues at the divide between metaphysics and epistemology. Amongst them, we should mention David Chalmers and Frank Jackson (Chalmers 1996; Jackson 1998). Amongst the critics we should mention Scott Soames (Soames 2007).

A recent approach to Frege’s puzzle that differs substantially from the canonical theories of the type we have considered so far is offered by Fiengo and May in their “De lingua beliefs” (Fiengo and May 2006). Their sophisticated treatment is based on the distinction between lexical items—that is, names such as “Tully” and “Cicero”—and the role these expressions play in “annotated” logical forms corresponding to the use of sentences containing name occurrences in concrete circumstances of communication and use. Their basic insight is that what refers is not the name as a lexical item but the name as part of a specific sentence occurrence. Moreover, the assignment of a semantic value to a given linguistic expression (say, a name) may be part of the interpretation of that very linguistic expression (these are the so-called *de dicto* readings). In general terms, the treatment of belief ascriptions and propositional attitude contexts is based on the full recognition that what is involved is not merely beliefs about the world, but also, and quite crucially, beliefs about language and grammar, crucially including beliefs about which assignments are involved in concrete instances of language use. To briefly illustrate, the use of the names “Tully” and “Cicero” in “John believes that Cicero was a Roman” and “John believes that Tully was a Roman” is bound to involve logical forms in which

these names are associated with different indexes—say, “Tully<sub>1</sub>” and “Cicero<sub>2</sub>”. Different indexes reflect de dicto beliefs according to which the two linguistic expressions are not necessarily associated with the same semantic value. In this way, one can derive the conclusion that the beliefs expressed in the two sentences are not the same belief, though the names “Tully” and “Cicero” are objectively associated to the very same semantic value. The framework developed by Fiengo and May applies to another puzzle, known as “Paderewski’s puzzle” and originally due to Kripke (Kripke 1979). The puzzle concerns contexts where two distinct occurrences of the very same name are involved. Even though the name “Paderewski” uniformly refers to an individual who was both a Polish statesman and a musician, there is no contradiction between John believing that Paderewski, as a musician, was musically talented and John believing that Paderewski, as a statesman, was not musically talented. In Fiengo and May’s framework, Paderewski’s puzzle is also explained on the basis of the indexes assigned to the name. On these grounds, Fiengo and May’s treatment of sentences like “Hesperus is Phosphorus” is that they are de re under their non-informative usage, whereby the same index is used for both names, and de dicto in their informative usage, whereby different indexes are used. When different indexes are used on “Hesperus” and ‘Phosphorus’, the identity statement involves both linguistic expressions and the object to which they refer, solving Frege’s puzzle. Fiengo and May’s analysis crucially relies on the semantic treatment of (co-)indexation they offered in their preceding book *Indices and Identity* (Fiengo and May 1994). In the semantic tradition within generative linguistics, this interpretation of indexes has been long and fiercely debated (for a different view on the semantics of indexes, see Heim and Kratzer 1998, which builds on Tanya Reinhart’s insights; Reinhart 1983a, b, 2000, 2006).

## References

- Bach, K. (1981). What's in a name? *Australasian Journal of Philosophy*, 59, 371–386.
- Burge, T. (1973). Reference and proper names. *Journal of Philosophy*, 70(14), 425–439.
- Chalmers, D. (1996). *The conscious mind*. New York/Oxford: Oxford University Press.
- Evans, G. (1973). A causal theory of names. *Proceedings of the Aristotelian Society*, 47, 187–208.
- Fiengo, R., & May, R. (1994). *Indices and identity*. Cambridge, MA: MIT Press.
- Fiengo, R., & May, R. (2006). *De Lingua belief*. Cambridge, MA: MIT Press.
- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100(1), 25–50.
- Haas-Spohn, U. (1995). *Versteckte Indexikalität und subjektive Bedeutung*. Berlin: Akademie Verlag.
- Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Oxford: Blackwell.
- Jackson, F. (1998). *From metaphysics to ethics*. Oxford: Clarendon Press.
- Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). Oxford: Oxford University Press.
- Kneale, W. C. (1962). Modality de dicto and de re. In E. Nagel, P. Suppes, & A. Tarski (Eds.), *Logic, methodology and philosophy of science, proceedings of the 1960 international congress* (pp. 622–633). Stanford: Stanford University Press.

- Kripke, S. A. (1979). A puzzle about belief. In A. Margalit (Ed.), *Meaning and use* (pp. 239–283). Dordrecht: Springer.
- Kripke, S. A. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.
- Reinhart, T. (1983a). *Anaphora and semantic interpretation*. Chicago: University of Chicago Press.
- Reinhart, T. (1983b). Coreference and bound anaphora: a restatement of the anaphora questions. *Linguistics and Philosophy*, 6, 47–88.
- Reinhart, T. (2000). Strategies of anaphora resolution. In H. Bennis, M. Everaert, & E. Reuland (Eds.), *Interface strategies* (pp. 295–324). Amsterdam: Royal Netherlands Academy of Arts and Sciences.
- Reinhart, T. (2006). *Interface strategies: Optimal and costly computations*. Cambridge, MA: MIT Press.
- Russell, B. (1905). On denoting. *Mind*, 14(56), 479–493.
- Searle, J. (1958). Proper names. *Mind*, 67(266), 166–173.
- Soames, S. (2007). *Reference and description. The case against two-dimensionalism*. Princeton: Princeton University Press.
- Strawson, P. F. (1959). *Individuals: An essay on descriptive metaphysics*. London: Methuen.

## Chapter 24

# Meaning and Acquaintance



### De Re and De Se

Our discussion of the first-person pronoun began with the observation that there is more to its meaning than just its reference. In Chap. 21, we mentioned two arguments in favor of this conclusion. The first is that the reference of the first-person, as well as that of indexicals in general, changes all the time, depending on its context of use. Hence, if we identify its meaning solely with its reference, we are left with no explanation of what is common to the different referents it picks up in the different contexts. The distinction between content and character, which we introduced in the previous chapters, addresses this problem. What is common to all the different uses of “I” is not its reference—or, to use Kaplan’s terminology, its content—but its character—the function that determines, for any context of use, what its content is going to be.

The second argument is that the first-person does not only express reference to the speaker in its context of use, it also expresses the speaker’s awareness that she is referring to herself. When Kazimir says “I”, we do not only understand that he is referring to the entity in the context of utterance that happens to be the speaker. We also understand that Kazimir is aware of the identity between the object he is referring to and himself. We proved this point by observing that it is perfectly possible for one to refer to oneself without being aware of doing so. This, for example, is the case with amnesiac Kazimir, who finds himself admiring the work of a certain painter, without being aware that that painter is he himself. Notably, the first-person pronoun is incompatible with such scenarios. When Kazimir says “I”, we understand that Kazimir is acknowledging the object he is talking about as him himself.

The challenge posed by this second problem is that of capturing the difference between de re propositions and what Lewis calls de se propositions. Consider sentences (1) and (2).

- (1) I am a painter
- (2) Kazimir is a painter

As we know, when uttered by Kazimir, the two sentences express the same content—the de re proposition that Kazimir is a painter. This is a de re proposition because it attributes a property—that of being a painter—to a *res*—the individual Kazimir. The imaginary scenario of amnesiac Kazimir shows us that the two sentences are, however, not completely equivalent. Although both (1) and (2) inform us that Kazimir is a painter, only (1) informs us that Kazimir is aware of the fact that the *res* he is referring to is him himself. In Lewis's terms, only (1) expresses a de se proposition—that is, a de re proposition where the *res* referred to is acknowledged by the speaker as him himself. What is, then, at the source of this difference?

We shall observe that the notion of character, at least in the form in which Kaplan conceived of it and which we have adopted in the previous chapters, is of no help in capturing the difference between de re and de se propositions. As we saw in the previous chapters, the character of an indexical is the function that identifies the reference of the indexical as the object in the context that meets a certain property. The character of “I”, for example, is the function that identifies the object in the context of use of “I” that is the speaker and assigns it to “I” as its reference—in Kaplan's terms, its content. The imaginary scenario of amnesiac Kazimir lost in the library, which we discussed in Chap. 21, shows us that the capacity cognitive subjects enjoy of identifying themselves as such cannot be expressed in the logic of objects and properties. People do not recognize themselves as such by coming to know a de re proposition about themselves. That is, they do not identify themselves as such by reasoning, “Look! That person is such and such, therefore, it must be me”. Such strategy may work in some cases, for example, when we recognize ourselves in an old picture because no one else could have been wearing such a silly shirt. However, it is a strategy that offers no logical guarantee. Recognizing a property in an individual is never a *sufficient* condition for one to acknowledge that individual as oneself. In principle, no property can provide such guarantee. As amnesiac Kazimir in the library, we could be aware of all the properties that are true of ourselves and still fail to recognize ourselves as such. If this observation is correct, we must conclude that there is no property we can use to define the character of “I” that would guarantee us, as a matter of logical necessity, that the individual who utters it is referring to herself in full awareness of doing so.

The character of “I”, as it is formulated by Kaplan, is the function that maps the *material* conditions in which “I” is used onto its reference. So formulated, this function tells us nothing about what the *cognitive* conditions are that enable speakers to say “I”. The problem, as we will see, has important implications for our understanding of the logical structure at the basis of linguistic meaning.



## The Beholder's Share

Long before the thought experiments of Lewis and Perry, Frege had already observed that the way individuals acknowledge themselves is “special and primitive”, in the sense that it cannot be reduced to any other existing logical notion. In Chap. 21, we referred to this mode of self-identification as *self-acquaintance*. To address the issue of de se propositions, we must enrich our model with this novel “special and primitive” notion.

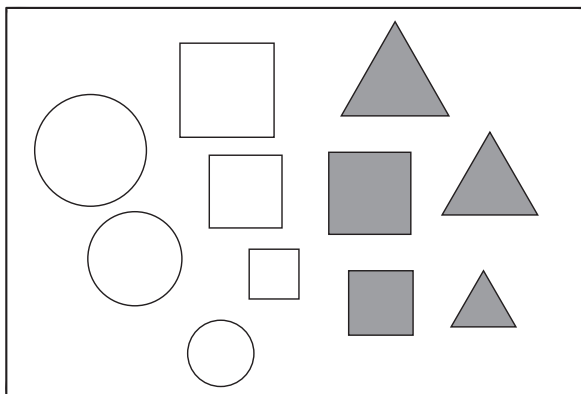
To address our challenge, we must, in a way, do what Gombrich did for perspective in the visual arts. In Gombrich's view, as we saw in Chap. 22, visual perspective emerges from the combination of two elements: on the one hand, the geometric principles that translate the three dimensions of the object that is depicted into the two dimensions of the image that depicts it; on the other, the perceptual capacities of the individual observing the image—that is, the cognitive principles that allow the observer to interpret the flat two-dimensional surface of an image as the depiction of a three-dimensional object. Both elements—observer and observed—are necessary for a comprehensive account of the notion of perspective. Perspective is not to be found exclusively in the object that is observed, but also in the eyes of the observer.

The problem we have identified with the first-person is of a similar nature. A comprehensive account of the meaning of “I” demands that we pay attention not only to the relation that binds its content to the material conditions of its use but also to the relation that binds its content to the cognition of the agents who use it. Kaplan's notion of character is insufficient in this respect. It provides a formalization of the relation between “I” and its material context but tells us nothing about the cognitive conditions that enable speakers to use it. To capture these conditions, we must add a psychological dimension to the notion of character. As Gombrich famously reminded us of the “beholder's share” in visual perspective, so we must now acknowledge the beholder's share in language.

## Acquaintance and Description

What is, then, self-acquaintance? This question has been debated by philosophers and cognitive scientists for a long time. A common position among many linguists, especially those interested in the semantics of de se propositions, is that self-acquaintance is a type of the more general notion of *acquaintance*. In the chapters to come, we will also present valuable reasons to disagree with this view. For the time being, however, it is useful to familiarize ourselves with it in some detail because it is a view that has a number of practical advantages, some of which we will explore in this and the next two chapters.

**Fig. 24.1** A model of geometric figures distinguished by shape—circle, triangle, and square—and color—white and grey



In his theory of knowledge, Russell famously distinguishes between two modes of gathering knowledge about the world around us: *acquaintance* and *description*. Knowledge by acquaintance is the knowledge we obtain by being directly aware of something through our senses. Knowledge by description is the knowledge we achieve through the intermediary of reasoning, by classifying the objects we are acquainted with on the basis of their properties. Consider, as an example, the model of geometric figures we used back in Chap. 6, here reproduced in Fig. 24.1.

There are two ways in which we can obtain knowledge of this picture. A first source of knowledge is mere observation. By setting our eyes on the picture, we see the different lines, shapes, and colors as well as their position in space. This is the knowledge we obtain from the picture by being visually acquainted with it. A second source of knowledge consists in recognizing that the objects we see share common properties. This is the process that provides us with the knowledge that, for example, some of the figures we see share common properties such as the property of being a circle, square, or triangle or that of being white or grey. This is knowledge we obtain by description.

Acquaintance and description differ in a number of respects. To begin with, acquaintance is *direct*, whereas description is *indirect*. Knowledge by acquaintance is the result of the direct causal effect that the external world produces on our senses. Knowledge by description, in contrast, is the result of an indirect process of inference. In its most basic form, it is the result of attributing properties to objects—such as when we classify a figure as being a circle or as being grey. We say that this sort of knowledge is inferential not only because it is the result of a deliberate mental process of classification, but also because it can be extended, by inference, to higher orders of logical complexity. For example, once we have recognized all the properties of all the objects in the picture above, we can also infer the relations that hold between them. For instance, we can infer that all circles are white and, therefore, that the property of being a circle logically entails, in the context of the picture, that of being white. By attributing properties to objects, we can then attribute properties to properties, properties to properties of properties, and so forth, and, in this way, produce knowledge of higher and higher logical complexity.

A further difference between knowledge by acquaintance and knowledge by description is that knowledge by acquaintance is *passive*, whereas knowledge by description is *active*. Acquaintance is the passive process of registering external information through the senses. As such, it does not involve any deliberate activity on the part of the knower. In fact, acquaintance is also, and necessarily, *concrete*. We can only be acquainted with things that exist materially and, therefore, can have a causal impact on our senses. Description, in contrast, is the active effort performed by the knower in organizing the information collected by attributing properties to its constituents.

A common way of expressing the distinction between knowledge by acquaintance and knowledge by description is to say that knowledge by acquaintance is knowledge of *what* is the case, whereas knowledge by description is knowledge *that* something is the case. In our example, knowledge by acquaintance is knowledge of what the figures are in the picture. Knowledge by description is knowledge that the figures are circles, squares, triangles, white, grey, and so forth. Hence, whereas knowledge by acquaintance provides us knowledge of objects—what is out there—knowledge by description provides us with knowledge of propositions—what properties are true of the objects out there.

The distinction between acquaintance and description we have illustrated here is due to Russell, although it finds its roots in the view of sensorial perception of René Descartes. In part III, we will see that there are serious reasons to question the validity of this view. It is true that our knowledge of the external world is based on two elementary capacities—that of recognizing objects and that of attributing properties to them. However, the research that has been performed in the last decades in the physiology, neurobiology, and psychology of perception provides a very different picture of how perception actually works. Far from what maintained by Russell and Descartes, sensorial perception has little to do with the passive, inert, unmediated, non-inferential registration of external stimuli that we described here.

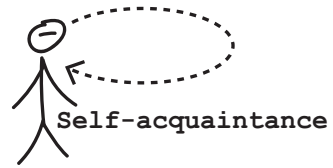
This said, for the time being we shall notice that the notion of acquaintance, as formulated by Russell, finds a number of useful applications in the context of language, starting with indexicals.

## Acquaintance and Self-Acquaintance

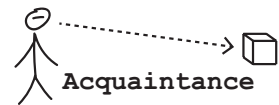
To begin with, Russell's notion of acquaintance allows us to formulate a working definition of *self-acquaintance*, which is, as we saw, an essential ingredient of the meaning of the first-person and the other indexicals. Acquaintance, as we saw, is the most basic form of informative relation that an individual can entertain with her environment (Fig. 24.2).

Amongst the objects one can be acquainted with in one's environment there is, of course, oneself. We will then maintain, for the time being, that self-acquaintance is simply acquaintance with oneself. The only difference between acquaintance and

**Fig. 24.3** The relation of self-acquaintance conceived as a relation of acquaintance one has with oneself



**Fig. 24.2** The relation of acquaintance is the most immediate informative relation an individual has with her environment



self-acquaintance is that, in the case of the former, perceiver and perceived are two distinct objects, whereas, in the case of the latter, perceiver and perceived are one and the same object (Fig. 24.3).

## Character Revised

Now that we have a working definition of self-acquaintance, we can use it to extend our account of the first-person to the problem at hand. So far, we have maintained that the character of “I” is the function that maps the context of its use into the individual that is currently speaking in that context. This definition, as we said, fails to capture the cognitive relation that the speaker holds with herself whenever referring to herself in the first-person. To capture this feature of the meaning of “I”, we will now adopt a richer notion of character. We will maintain that the character of “I” is the function that maps the context of its use into *the individual that the speaker is acquainted with as herself* in that context. According to this definition, the character of “I” now encompasses two distinct dimensions. The first is its material dimension, which describes the causal link between the context of use of “I” and its reference. The second is its epistemic dimension, which describes the cognitive link between the speaker uttering “I” and its reference. The first dimension ensures that the content of “I” is the individual speaking in the context of utterance, the second that the content of “I” is the individual the speaker is acquainted with as herself in the same context.

This enriched notion of the character of “I” allows us to finally explain the specific properties of de se propositions. According to our new analysis of the character of “I”, speakers are able to refer to themselves as “I” only when satisfying the cognitive condition of being acquainted with themselves as such. Hence, when Kazimir says “I”, we not only infer that he is talking about himself but also that he is aware that the object of his reference is him himself.

In the following two chapters, we will extend this framework to the domain of propositional attitudes. This will give us a chance to explore the potential of the ideas we have introduced in this chapter but also to identify an element of epistemic subjectivity within the interpretative structure of natural language meaning. In Chap. 28, however, we will also consider some critical arguments against the view of self-acquaintance we have introduced in this chapter.

## References and Remarks

In one of the last sections of his essay “Demonstratives” (Kaplan 1989, section XVII “Epistemological Remarks”), Kaplan discusses the possibility that the character of indexicals is all we need to capture what he refers to as their “cognitive significance”. Kaplan admits with Frege that there is something special and primitive about the way individuals are presented to themselves but also finds that the notion of character is all we need to capture this observation within the limits that are relevant to the formulation of a semantic theory of the first-person and the other indexicals. Whether characters offer sufficient logical means to represent the cognitive significance of indexicals remains an issue of debate. Criticism of the epistemological implications of Kaplan’s formalization can be found in Wettstein (1986), Taschek (1987), and Crimmins (1992). Some scholars have proposed to identify the cognitive significance of indexicals with the notion of a ‘diagonal proposition’ (see Stalnaker 1978; Haas-Spohn 1995). The criticism we have discussed in this chapter is based on a more recent essay by Kripke (Kripke 2011) where he criticizes Kaplan’s semantics of the first person by observing that, for Kaplan’s notion of character to do the job, it must presuppose that speakers are already endowed with a primitive notion of self. His conclusion is that “each of us does have a special acquaintanceship with himself or herself, as philosophers from Descartes to Frege have held. This self-acquaintance is more fundamental than anything purely linguistic, and is the basis of our use of first person locutions. And each of us can use them to make genuine claims, to express genuine propositions” (Kripke 2011, p. 319). Philosophers have been wary of this conclusion because, as Kripke’s quote suggests, it pulls in the direction of a Cartesian view of the self whereby the self is an immaterial, disembodied primitive whose essence cannot be reduced to any of the material or abstract properties of the external world. We will return to this issue in Chaps. 27, 28, and 29.

Russell discusses the distinction between knowledge by acquaintance and knowledge by description first in his article “Knowledge by Acquaintance and Knowledge by Description” (Russell 1910) and, then, in his book *The Problems of Philosophy* (Russell 1912; in particular, Chap. 5) and in his article “On the Nature of Acquaintance” (Russell 1914). Interestingly, Russell was doubtful about the possibility of conflating the notion of self-acquaintance with that of acquaintance (see especially Russell 1914). Needless to say, the issue has been—and, in fact, still is—a topic of heated philosophical debate. What is interesting to observe, from our

perspective, is that the treatment of self-acquaintance as a form of acquaintance has been taken for granted by a number of linguists, particularly those who have attempted at providing accounts of *de se* that reduce it to a special case of *de re*. Notable examples are the frameworks proposed by Lewis (Lewis 1979), Maxwell Cresswell and Arnim von Stechow (Cresswell and von Stechow 1982) and, more recently, Emar Maier (Maier 2009). We will review the advantages of this type of approach in the following two chapters. In Chap. 27, however, we will also discuss a form of *de se* that cannot be captured by the notion of self-acquaintance, as we have formulated it in this chapter. In part III we will extend our criticism to the notion of acquaintance as a whole (see, in particular, Chap. 34).

## References

- Cresswell, M. J., & von Stechow, A. (1982). *De re* belief generalized. *Linguistics and Philosophy*, 5(42), 503–535.
- Crimmins, M. (1992). *Talk about beliefs*. Cambridge, MA: MIT Press.
- Haas-Spohn, U. (1995). *Versteckte Indexikalität und subjektive Bedeutung*. Berlin: Akademie Verlag.
- Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). Oxford: Oxford University Press.
- Kripke, S. A. (2011). The first person. In S. A. Kripke (Ed.), *Philosophical troubles* (pp. 292–321). Oxford: Oxford University Press.
- Lewis, D. (1979). Attitudes *de dicto* and *de se*. *Philosophical Review*, 88, 513–543.
- Maier, E. (2009). Presupposing acquaintance: A unified semantics for *de dicto*, *de re* and *de se* belief reports. *Linguistics and Philosophy*, 32(5), 429–474.
- Russell, B. (1910). Knowledge by acquaintance and knowledge by description. *Proceedings of the Aristotelian Society*, 11, 108–128.
- Russell, B. (1912). *The problems of philosophy*. Oxford: Oxford University Press.
- Russell, B. (1914). On the nature of acquaintance. *Monist*, 24, 161–187.
- Stalnaker, R. C. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics vol. 9: Pragmatics* (pp. 315–332). New York: Academic Press.
- Taschek, W. (1987). Content, character, and cognitive significance. *Philosophical Studies*, 52, 161–189.
- Wettstein, H. (1986). Has semantics rested on a mistake? *Journal of Philosophy*, 83, 185–209.

## Chapter 25

# Attitudes De Se



The study of indexicals demonstrates that the function of language is not that of merely describing a model, but, in fact, that of describing it from the vantage point of a perspective. In the previous chapters, we saw that such a perspective is actually twofold—a material one, representing the material coordinates of the context, and a psychological one, representing the cognitive dimension of the observer. In particular, in the previous chapter we have considered the possibility of explaining this cognitive dimension in terms of Russell’s notion of acquaintance. In this and the following chapter, we will see that this understanding of the notion of perspective extends to other domains of language.

As we saw in Chap. 13, propositions are not only the referents of declarative sentences. They are also the objects of propositional attitudes—the things people believe, know, desire, expect, hope, remember, imagine, and so on. We shall see now that, as we express propositions from a perspective, so we entertain beliefs, desires, memories, and expectations. In particular, we will see in this chapter that, as natural language provides the means to unambiguously express *de se* propositions, so it provides the means to unambiguously express *de se propositional attitudes*.

### Attitudes De Re and Attitudes De Se

The finding that natural language allows speakers to unambiguously express *de se* propositional attitudes is a relatively recent one. Part of the reason why this is so is that the immediate evidence seems to suggest otherwise. Suppose that Frida wants to report the fact that Kazimir uttered sentence (1) by using an *indirect* report. Most likely, she would say something like (2).

- (1) I am a painter
- (2) Kazimir said that he is a painter

Notice that, in order to produce an indirect report of Kazimir's utterance of (1), Frida must make a substantial change to Kazimir's original words: She must replace the first-person pronoun "I" of Kazimir's original utterance with the third-person pronoun "he". The reason for this change is obvious. If Frida were to maintain the first-person pronoun of Kazimir's original utterance, she would end up saying (3).

- (3) Kazimir said that I am a painter

Sentence (3), however, reports something different from what Frida intends to report. Uttered by Frida, (3) says that Kazimir said that *Frida* is a painter. This is not what Frida wants to say. What she really wants to say is that Kazimir said that Kazimir himself is a painter. Sentence (3) fails to report Kazimir's words faithfully because "I" is bound to change its meaning from Kazimir to Frida whenever uttered by Frida. Hence, if Frida wants to report something Kazimir said about himself, she must not use the first-person pronoun of Kazimir's initial utterance, as in (3), because that is bound to refer to her whenever in her mouth. To report the fact that Kazimir said (1), Frida must then resort to a third-person pronoun, as in (2). However, replacing the first-person pronoun of Kazimir's original utterance with a third-person expression comes with significant costs. Compared to Kazimir's original sentence, the report made by sentence (2) is ambiguous in two ways.

The first ambiguity affecting (2) is an ambiguity of content. Consider first sentence (1). Whenever uttered by Kazimir, the sentence expresses its content without ambiguities. This is the proposition that Kazimir is a painter. Once the material circumstances under which the sentence is uttered are known, the reference of "I" can be individuated unambiguously: It is the person speaking—that is, Kazimir. This, however, is not the case with "he". Like "I", "he" is an indexical, in the sense that its meaning is not established lexically but is chosen amongst the objects made available by its context of use. However, "he" differs from "I" in that it does not come with a well-defined character like "I"—a function that selects, univocally and without ambiguity, an object of the context. All an object has to do to qualify as a potential referent of "he" is to be singular, male-gendered, and animate. In fact, the reference of "he" is not really determined by the context, as in the case of "I". It is, rather, *inferred* from it on the basis of other contextual clues. It can be the person the speaker is pointing at, the main topic of the current conversation, the referent of a name that has just been uttered. In the case of sentence (2), we tend to understand "he" as referring to Kazimir, because his name is mentioned immediately before, as the subject of the main clause. However, this is not necessarily so. Suppose, for example, that Frida wanted to report the fact that Kazimir uttered sentence (4), rather than sentence (1).

- (4) Pablo is a painter

To report the fact that Kazimir uttered (4), Frida could still use sentence (2), as long as she makes clear that, in the context in which she is uttering it, she intends



“he” to refer to Pablo, not Kazimir. The conclusion is that, when Frida utters sentence (2) to report Kazimir’s utterance, she loses the specific quality of Kazimir’s original utterance that makes it unambiguously about Kazimir in the context of his talking. Whereas sentence (1), whenever uttered by Kazimir, is bound to express the proposition that Kazimir is a painter, sentence (2), whenever uttered by Frida, remains free to express different propositions depending on the reference that is chosen by the interpreter as the reference of “he”.

The second ambiguity that affects (2) concerns its character. As we have learned, sentence (1) expresses not only a *de re* proposition, but also a *de se* one, whereby the speaker is aware that he is referring to himself. This is not the case with sentence (2). Even when we have correctly inferred that Frida’s intention in using “he” is to refer to Kazimir and, therefore, to report something Kazimir said about he himself, we cannot determine from (2) alone whether, in his original utterance, Kazimir used the first-person pronoun to refer to himself. Hence, we cannot determine whether Kazimir had any awareness that he was talking about he himself. To appreciate this observation, let us make use again of amnesiac Kazimir. Suppose amnesiac Kazimir is still at the museum in a profound state of amnesia and, having reviewed his own paintings without awareness that they are his own, utters (5).

(5) Kazimir is a painter

In this scenario, Kazimir is making a statement about himself without being aware of doing so. The proposition he expresses by uttering (5) is *de re*, but not *de se*. It is, that is, a proposition Kazimir expresses about Kazimir without awareness of the identity between himself and the object the proposition is about.

Notably, even in this scenario, Frida can still report Kazimir’s utterance of (5) by using (2). Sentence (2), that is, can function perfectly well as a truthful report that Kazimir said of himself that he is a painter while not being aware of doing so. All that is required for sentence (2) to be true is that Kazimir said of Kazimir that he is a painter. Whether Kazimir himself was or was not aware that he was talking about himself is irrelevant when establishing the truth of (2). We find that, whereas (1) unambiguously expresses a *de se* proposition, (2) does not unambiguously express a *de se* propositional attitude. This is not surprising. The moment the first-person pronoun “I” of Kazimir’s original utterance is replaced with the third-person pronoun “he”, the character of Kazimir’s original utterance is also lost and, with it, the relation of self-acquaintance it expresses.

The conclusion we seem to reach from these observations is that the distinction between *de re* and *de se* is lost in indirect reports. This is not only true for the verb “say”. It also applies to all other verbs of propositional attitudes. Consider the sentences in (6).

(6a) Kazimir thinks that he is a painter

(6b) Kazimir believes that he is a painter

Neither of these sentences tell us whether Kazimir’s reported thought or belief is *de re* or *de se*. We can infer from these sentences that Kazimir has a certain thought or belief about himself. But we cannot as legitimately conclude that Kazimir is

aware of being the object of his own thought or belief. The general conclusion seems, then, that there is no distinction between *de re* and *de se* in the context of indirect reports of propositional attitudes. The reason seems obvious enough. When an utterance, thought, belief, or other attitude is reported in someone else's words, all its original indexical expressions must be removed. As a result, the perspective under which the original utterance, thought, or belief was originally entertained is lost. Verbs of propositional attitude succeed in reporting the content of propositional attitudes, but fail to preserve their character.

## Value and Logic of Attitudes De Se

The negative conclusion we seem to reach is that the distinction between *de re* and *de se* propositions is lost when propositions become the objects of verbs of propositional attitude. This conclusion is surprising, for at least two reasons. To begin with, it is surprising to find that language cannot express the distinction between *de re* and *de se* attitudes when the distinction plays such a central role in the cognitive life of individuals. It does make a significant difference to us—cognizing agents—to think, believe, know, hope, or regret propositions that are *de re* as opposed to propositions that are *de se*. There is an important difference between believing that somebody is in danger and believing that *we* are in danger as there is an important difference between knowing that somebody won the lottery and knowing that *we* won the lottery. Human beings are able to locate themselves in their environment not only when talking, but also when forming beliefs, thoughts, expectations, and other attitudes. Propositional attitudes that are eminently *de se* represent the capacity to navigate the environment from the first-personal perspective of a cognizing agent. It is surprising to find that natural language lacks the means to express this simple yet fundamental cognitive capacity.

The second reason for doubting the negative conclusion reached above is that the possibility of having indirect reports of *de se* propositional attitudes is perfectly compatible with the logical framework we have developed in the previous chapters. In Chap. 13, we have argued with Hintikka that verbs of propositional attitude denote relations between individuals and sets of possible worlds—that is, contents. It is perfectly legitimate to conceive of an extension of such framework where verbs of propositional attitude are relations between individuals and richer propositional structures expressing contents along with their characters. The logical framework we have developed to explain indexicals offers, in principle, the logical structure to unambiguously express the content of propositional attitudes without losing their character.

So, why do we not find unambiguous expressions of *de se* propositional attitudes across natural languages?

## The Grammar of Attitudes De Se

Of course, there is a simple and straightforward way to retain the character of someone else's propositional attitude: making a *direct* report. Compare (2), repeated below, to (7).

- (2) Kazimir said that he is a painter  
 (7) Kazimir said, "I am a painter"

As we saw, the indirect report in (2) is ambiguous in two ways. It is ambiguous because "he" can refer to different individuals in different contexts and because, even when the context ensures that "he" refers to Kazimir, the sentence fails to express whether the proposition Kazimir originally entertained was *de re* or *de se*. In contrast, the direct report in (7) is unambiguous. It clearly reports a proposition Kazimir entertained about himself and it unambiguously tell us that Kazimir original proposition was a *de se* proposition, expressed in the first-person.

In the last decades, linguists have uncovered linguistic evidence that *de se* propositional attitudes can also be expressed by indirect reports. Suppose we wanted to translate sentence (2) in Italian. There are two ways of doing so. The first is (8):

- (8) *Kazimir ha detto che lui è un pittore*  
 Kazimir has said that he is a painter

Sentence (8) is by and large a word-by-word translation of its English counterpart. In fact, the sentence is ambiguous in the same way as (2) is. Similarly to "he" of sentence (2), the pronoun "lui" of (8) is free to refer to any animate, singular, male gendered entity that is relevant in the context in which the pronoun is used. Sentence (8) is also similar to (2) in that it is ambiguous between reporting a *de re* and a *de se* propositional attitude. Even in a context that makes it clear that "lui" is intended to refer to Kazimir, we cannot infer from (8) whether Kazimir was aware of referring to himself when making his original utterance. There is, however, a second way to translate (2) in Italian:

- (9) *Kazimir ha detto di essere un pittore*  
 Kazimir has said to be a painter

Grammatically, (9) differs from (8) in three respects. Firstly, in place of the complementizer "che"—English "that"—we find the preposition "di". Secondly, the embedded verb "essere" is in the infinitival form. Finally, the embedded sentence "di essere un pittore" has no overt subject—that is, compared to (8), the pronoun "lui" is omitted. Sentence (9) differs from sentence (8) also in its meaning. Whereas (8) is ambiguous in the same way as its English counterpart (2), sentence (9) is not. To begin with, (9) can only be interpreted as reporting a *de re* belief that Kazimir holds about Kazimir himself. It cannot report a belief Kazimir holds towards a different *res*, as is the case with (8). It is for this specific reason that grammatical structures such as the one exemplified by (9) are referred to by linguists as *control* structures. The label applies to those structures where the subject of the

embedded clause is omitted and is necessarily understood as equivalent to the subject of the main clause. Sentence (9) qualifies as a control structure because the subject of the embedded verb “essere” is omitted but is automatically understood as Kazimir, the subject of the matrix clause, and cannot be understood as anyone else. Sentence (9) differs from sentence (8) also in that it unequivocally reports a de se proposition. As it was first observed by the linguists Gennaro Chierchia and Jerry Morgan, control structures always and unambiguously express de se propositional attitudes. Whereas (8) does not allow us to infer whether Kazimir’s original utterance was a de re or de se proposition, (9) unambiguously expresses the fact that Kazimir said something about himself in full awareness of doing so.

Notably, control structures such as the one exemplified in (9) can be found in a variety of languages. In fact, we find them also in English, although not with the verb “say”. We can observe a control structure of the same type as (9) with the verb “expect”. Suppose we want to report Kazimir’s expectation that he will win the lottery. There are two ways to go about it in English—(10) or (11).

(10) Kazimir expects that he will win the lottery

(11) Kazimir expects to win the lottery

Sentences (10) and (11) differ in their meaning in exactly the same way as the Italian sentences (8) and (9) do. Sentence (10) may express an expectation Kazimir holds towards himself as well as one he holds towards someone else, depending on how we interpret “he”. Conversely, (11) necessarily expresses an expectation Kazimir holds towards himself. Furthermore, whereas (10) is ambiguous between reporting a de re and a de se propositional attitude— even when “he” is interpreted as referring to Kazimir—(11) is always and necessarily interpreted as expressing a de se expectation, which Kazimir holds towards himself in full awareness of doing so.

In contrast to our initial negative conclusions, control structures offer a grammatical format, shared by a number of natural languages, to produce indirect reports of propositional attitudes that are unambiguously de se.

## Logophoricity

Control structures are not the only class of natural language expressions that unambiguously express de se propositional attitudes. At least another class deserves to be discussed—*logophoric pronouns*. Logophoric pronouns were first observed in languages of the Niger-Congo family by the linguist George Clements. They are typically characterized by two main properties. The first is that they can only occur in embedded sentences that are the complement of a verb of propositional attitude. The second is that they can refer exclusively to the subjects of such propositional attitudes. Consider, as an example, the case of Ewe’s logophoric pronoun “yè”, famously discussed by Clements. Compare the Ewe’s sentences (12) and (13):

- (12) *Kofi be e-dzo*  
Kofi say he-leave
- (13) *Kofi be yè-dzo*  
Kofi say LOG-leave

Consider first sentence (12) together with its word-by-word translation. “Kofi” is a proper name and the subject of the matrix sentence. “Be” is the main verb, corresponding to English “say”. “E-” is a pronoun, corresponding to English “he”. Finally, “dzo” is the embedded verb, translated in English as “leave”. The whole sentence translates in English as “Kofi said he left”. As in its English translation, the pronoun “e-” may pick different referents in different contexts. Hence, the pronoun can, but is not bound to, refer to Kofi. This is not the case with sentence (13). Sentence (13) is in all respects identical to (12), with the only exception that, in place of the pronoun “e-”, we find the logophoric pronouns “yè-”. This pronoun differs from “e-” in that it is bound to refer to the subject of the main sentence, Kofi. The logophoric pronoun performs a function equivalent to that performed by control structures in that it binds the interpretation of the subject of the embedded clause to that of the subject of the matrix clause.

Since Clements’ original findings, logophoric pronouns have been attested in a large variety of languages. Outside the Niger-Congo family, we find them in East Asian languages such as Chinese and Japanese, Germanic languages such as Icelandic, and Romance languages such as Italian. They are also attested in ancient Indo-European languages, such as Ancient Greek. English can construct a logophoric pronoun by combining a regular pronoun, such as “he”, with a reflexive pronoun, such as “himself”. Compared to (2), repeated below, sentence (14) can only express something that Kazimir said about himself.

- (3) Kazimir said that he is a painter
- (14) Kazimir said that he himself is a painter

Crucial for us is to observe that logophoric pronouns are typically unambiguously *de se*. That is, not only do they ensure that the subject of the propositional attitude and the *res* the propositional attitude is about are the same; they also unambiguously express awareness on the part of the subject that this is so. This goes for Ewe’s “yè-” as well as for English “he himself”. Sentence (14), for example, unambiguously reports the fact that Kazimir said something of himself while being fully aware of doing so.

Control and logophoric pronouns are two of the most significant sources of evidence that, contrary to our initial conclusions, natural language does have the means to produce indirect reports of propositional attitudes that are unambiguously *de se*. In the next chapter, we will provide a semantics for verbs of propositional attitudes that accounts for this possibility.

## References and Remarks

In his “Demonstratives” (Kaplan 1989), Kaplan makes a point that characters are an exclusively root clause phenomenon. The conceptual reason behind his point is simple. Characters are not the meaning of linguistic expressions; rather, they are the functions that assign meaning to linguistic expressions on the basis of the context in which they occur; hence, they are not allowed to participate in the compositional machinery of natural language meaning. Kaplan admits the theoretical possibility of defining logical operators manipulating characters, but also maintains that they do not occur in natural language. In fact, Kaplan calls such operator “monsters”. If Kaplan were correct, we should indeed expect that indirect reports cannot unambiguously express *de se* propositional attitudes.

To be sure, the logical possibility of unambiguously expressing *de se* propositional attitudes in indirect reports had already been contemplated by the philosopher Henri-Neri Castañeda in two fundamental articles—“‘He’: A study in the logic of self-consciousness” (Castañeda 1966) and “On the logic of attributions of self-knowledge to others” (Castañeda 1968). Although Castañeda does not use terms such as *de se*, which would be introduced later by Lewis, or logophoric pronoun, he conceives of the theoretical possibility of a pronoun that is, in effect, equivalent to a logophoric pronoun in that it occurs in the embedded clause of indirect reports, is necessarily bound to co-refer with the subject of the main clause, and unambiguously expresses a *de se* attitude towards the proposition expressed by the embedded clause.

The observation that control structures unambiguously express *de se* propositional attitudes is due to Jerry Morgan and Gennaro Chierchia (Morgan 1970; Chierchia 1989) although it was Chierchia who first provided an explicit compositional semantics of *de se* attitudes.

For a thorough and comprehensive overview of the different perspectives on *de se* phenomena in indirect reports, including many of the relevant philosophical, cognitive and pragmatic issues, the interested reader is referred to Alessandro Capone’s “The pragmatics of indirect reports” (Capone 2016).

Logophors have been known in the field of linguistic typology since Clements’ pioneering work on Ewe (Clements 1975) and have been observed in a number of different languages. In the field of semantics, it is commonly maintained that logophors unambiguously contribute *de se* propositional attitudes (see, for example, Schlenker 1999; Stephenson 2007a; von Stechow 2003). This assessment, however, has been recently put into question by Hazel Pearson (Pearson 2015). Pearson notices that the typological literature on logophors typically discusses their semantic properties only informally and only rarely explicitly addresses the issue of whether logophors are unambiguously *de se*. Hopefully, future research will offer more clarity on this topic.

Another phenomenon that is relevant to the issue of *de se* propositional attitudes and has recently attracted the attention of semanticists is that of “shifted

indexicals”. Shifted indexicals are indexicals that can be interpreted relatively to the context of an indirect report, rather than that of an actual speech act. In some languages, indexicals such as the first-person can be understood as referring to the subject of an indirect report. A language of this sort is Amharic. The Amharic equivalent of a sentence such as (4)—“Kazimir said that I am a painter”—is, in fact, ambiguous between two possible interpretations. The first, equivalent to the English one, is that Kazimir said that the individual uttering the sentence is a painter. The second, unattested in English, is that Kazimir said that he himself, as the speaker in the reported speech act, is a painter. Philippe Schlenker (Schlenker 1999, 2003) has further observed that these shifted uses of the first-person are unambiguously *de se*. These observations vigorously contradict Kaplan’s restriction against monsters (on this, see also Delfitto and Fiorin 2011, 2014).

Other domains of language where we find the unambiguous expression of *de se* propositions and propositional attitudes are predicates of personal taste (Pearson 2013; Stephenson 2007b), epistemic modals (Stephenson 2007b), and the narrative style known as free indirect speech (Delfitto et al. 2016).

## References

- Capone, A. (2016). *The pragmatics of indirect reports: Sociophilosophical considerations*. Heidelberg: Springer.
- Castañeda, H.-N. (1966). ‘He’: A study in the logic of self-consciousness. *Ratio*, 8, 130–157.
- Castañeda, H.-N. (1968). On the logic of attributions of self-knowledge to others. *Journal of Philosophy*, 65, 439–456.
- Chierchia, G. (1989). Anaphora and attitudes *de se*. In R. Bartsch, J. van Benthem, & P. van Emde Boas (Eds.), *Semantics and contextual expressions* (pp. 1–31). Dordrecht: Foris.
- Clements, G. N. (1975). The logophoric pronoun in Ewe: Its role in discourse. *Journal of West African Languages*, 10, 141–177.
- Delfitto, D., & Fiorin, G. (2011). Person features and pronominal anaphora. *Linguistic Inquiry*, 42(2), 193–224.
- Delfitto, D., & Fiorin, G. (2014). Indexicality and left-periphery. *Rivista di Linguistica*, 26(1), 59–92.
- Delfitto, D., Fiorin, G., & Reboul, A. (2016). The semantics of person and *de se* effects in free indirect discourse. *SpringerPlus.*, 5, 1451.
- Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). Oxford: Oxford University Press.
- Morgan, J. L. (1970). On the criterion of identity for noun phrase deletion. In *Papers from the sixth regional meeting of the CLS* (pp. 380–389). Chicago: The Chicago Linguistics Society.
- Pearson, H. (2013). A judge-free semantics for predicates of personal taste. *Journal of Semantics*, 30, 103–154.
- Pearson, H. (2015). The interpretation of the logophoric pronoun in ewe. *Natural Language Semantics*, 23(2), 77–118.
- Schlenker, P. (1999). Propositional attitudes and indexicality: A cross-categorical approach. PhD dissertation, MIT.
- Schlenker, P. (2003). A plea for monsters. *Linguistics and Philosophy*, 26, 29–120.

- Stephenson, T. C. (2007a). Towards a theory of subjective meaning. PhD dissertation, MIT.
- Stephenson, T. C. (2007b). Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy*, 30, 487–525.
- von Stechow, A. (2003). Feature deletion under semantic binding. In M. Kadowaki & S. Kawahara (Eds.), *Proceedings of NELS 33* (pp. 133–157). Amherst, MA: GLSA Publications.



# Chapter 26

## Worlds and Centers



### Limits of Hintikka's Semantics

Now that we have demonstrated the possibility of expressing de se propositional attitudes in natural language, we shall consider how our current model of interpretation of natural language can be extended to capture them. In this chapter, we will present an extension of Hintikka's semantics of verbs of propositional attitudes that does just that.

Let us first briefly return to Hintikka's theory, as we know it from Chap. 13. The essential idea of the theory is that verbs of propositional attitude such as "believe" contribute relations between individuals and propositions. A sentence of the form " $x$  believes that  $p$ " is true if, and only if, the proposition that describes what the subject  $x$  believes entails the proposition  $p$ . In the jargon of possible worlds, we say that the sentence is true if, and only if, the possible worlds that are consistent with what  $x$  believes to be the case are a subset of the possible worlds where  $p$  is also the case. We expressed these conditions as in (1), where  $BEL_x$  symbolizes the set of possible worlds that are consistent with what  $x$  believes to be the case—that is,  $x$ 's *modal base*—and the formula  $BEL_x \subseteq p$  expresses the condition that  $x$ 's modal base is located within the region of the logical space that is consistent with the proposition  $p$ .

- (1)  $p \rightarrow (x \rightarrow \text{the set of possible worlds where } BEL_x \subseteq p)$

It is important to observe that, in this format, Hintikka's semantics is not fine-grained enough to capture de se propositional attitudes. Compare (2) and (3).

- (2) Kazimir believes that he is a painter  
(3) Kazimir believes that he himself is a painter

As we saw in the previous chapter, sentence (2) is ambiguous between the report of a de re and the report of a de se propositional attitude. This is because the

pronoun “he” is ambiguous both in what it refers to and in how it refers to it—the relation of acquaintance it expresses. Sentence (3), conversely, is unambiguously *de se*. This is because the logophoric pronoun “he himself” is bound to refer to the subject of the matrix sentence—“Kazimir”—and to express Kazimir’s first-personal acquaintance with the object of his belief as he himself. Hintikka’s semantics, as it stands, cannot capture the specific *de se* quality of (3) and, therefore, cannot describe the difference between the two sentences. If, in fact, we adopt Hintikka’s analysis as it stands, all we can say about sentence (3) is that it is true if, and only if, all the possible worlds that are consistent with what Kazimir believes are also possible worlds where Kazimir is a painter. All these truth-conditions require for (3) to be true is that Kazimir’s belief is consistent with the proposition that the individual Kazimir is a painter. These truth-conditions tell us nothing about Kazimir’s acquaintance with the object of his belief.

In the terms of Kaplan’s distinction between content and character, we can formulate the problem as follows. In Hintikka’s semantics, verbs of propositional attitude express relations between an individual and a set of possible worlds—that is, a content. This relation is not fine-grained enough to capture the properties of the perspective that relates the subject to the object of the attitude. Hintikka’s theory, that is, is limited to regarding propositional attitudes as relations between an individual and a content when, in order to capture the specific perspective a subject has towards a content, what we need is a character.

The problem, then, is that Hintikka’s analysis fails to capture the difference in meaning between (2) and (3) and, more precisely, the distinctive perspective that Kazimir has towards the content of his belief according to (3). To be able to properly account for the specific semantic quality of *de se* attitudes, we must allow the notions of context and acquaintance to take part in the relation expressed by verbs of propositional attitudes. We must, in other words, allow characters to participate to the compositional machinery of natural language meaning.

## Centered Possible Worlds

So far, we have maintained that contexts are complementary to models, as exemplified in Fig. 26.1. They combine with models to provide super-models. Together, context and model provide a description of a state of affairs, along with the coordinates of the perspective from which such description is entertained.

**Fig. 26.1** A super-model made of a context and a model



The model of natural language meaning is also characterized, as we saw, by the dimension of possibility. It does not only depict the world as it *is*, but also as it *can* be. We visualized this type of model as made of a number of layers, each representing a possible arrangement of its constituting elements—that is, a *possible world* (Fig. 26.2).

Now, each possible world is, in and by itself, an individual description of a specific possible state of affairs which can be acknowledged from a specific possible perspective. In principle then, each individual possible world can be independently anchored to a context, representing the perspective from which it is contemplated (Fig. 26.3).

Then, as we have combined models with contexts to create super-models, we can now combine possible worlds with contexts to create super-possible worlds, although the literature adopts a different terminology. The context of a possible world is commonly referred to as its *center*. Correspondingly, super-possible worlds are referred to as *centered possible worlds*.

A useful metaphor to understand the notion of centered possible world is offered by Lewis: Whereas possible worlds are “maps” of the world, centered possible worlds are maps of the world that also have an arrow in them that says “you are here”. They are descriptions of a state of affairs that are also anchored to a certain context—that is, a perspective.

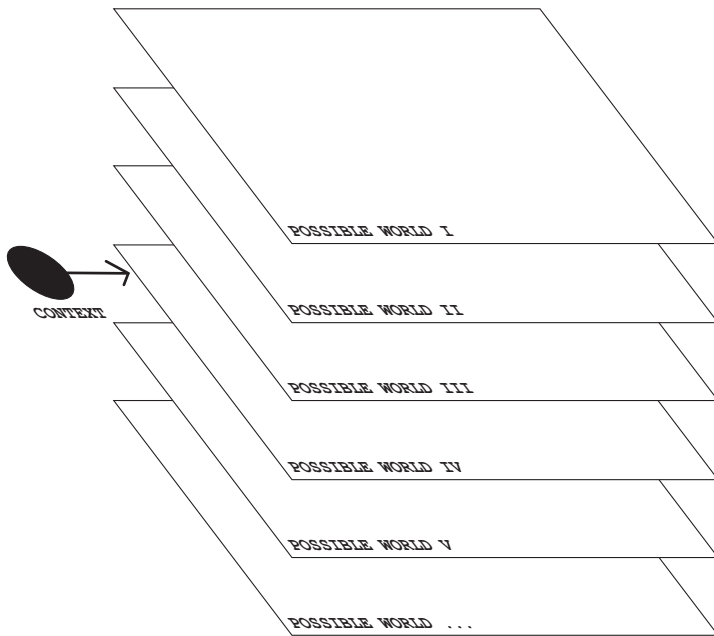


Fig. 26.2 A super-model made of a context and a collection of possible worlds

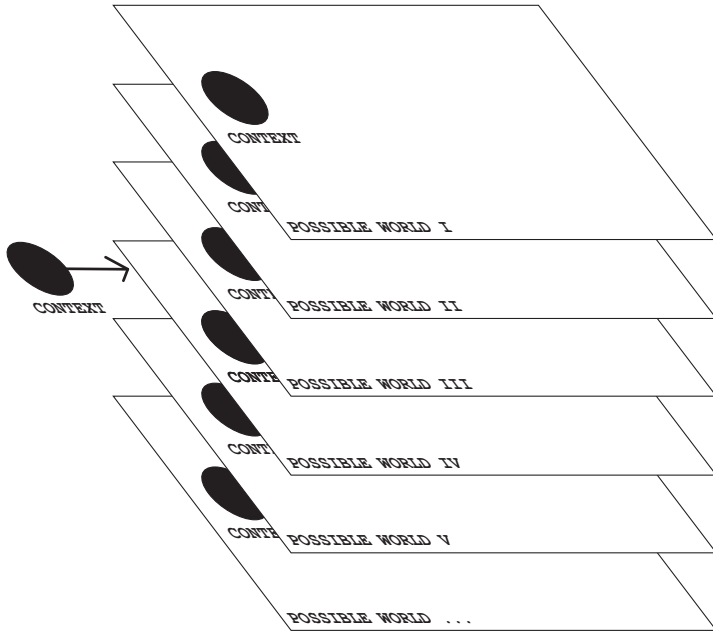


Fig. 26.3 A super-model made of a context and a collection of *centered* possible worlds

## Centered Propositions

Thanks to the notion of centered possible world, we can now provide a richer notion of proposition—that of a *centered proposition*. Since Chap. 11, we have taken propositions to correspond to sets of possible worlds—all the possible worlds that are consistent with a certain description. In the same spirit, we will now take centered propositions to be sets of centered possible worlds—all the possible worlds that are consistent with a certain description as well as a certain perspective. When expressing Hintikka’s semantics of the verb “believe”, we used the symbol  $p$  as a variable for general reference to propositions. We will now use the symbol  $p_c$  to refer to centered propositions. More exactly, we will take the symbol  $p_c$  to represent the set of possible worlds that are consistent with the proposition  $p$  as they are acknowledged from the center  $c$ .

## Hintikka’s Semantics Revised

Hintikka’s semantics of “believe”, repeated below, can be easily restated as a relation between individuals and centered propositions.

- (1)  $p \rightarrow (x \rightarrow \text{the set of possible worlds where } BEL_x \subseteq p)$

First, we must redefine the modal base  $BEL_x$  in the form of a centered proposition. Let  $BEL_x$  be the set of all *centered* possible worlds that are consistent with what  $x$  believes as they are acknowledged from  $x$ 's perspective.  $BEL_x$  is then the centered proposition that represents what  $x$  believes to be case from  $x$ 's own perspective. Hintikka's semantics can then be reformulated as follows:

- (4)  $p_c \rightarrow (x \rightarrow \text{the set of centered possible worlds where } BEL_x \subseteq p_c)$

According to (4), "believe" denotes a relation between a centered proposition  $p_c$  and an individual  $x$  which is true if, and only if, the centered proposition  $BEL_x$  is a subset of  $p_c$ . The relation is true if, and only if, all the possible worlds that are consistent with what  $x$  believes to be the case from  $x$ 's own perspective are also possible worlds that are consistent with proposition  $p$  as acknowledged from perspective  $c$ . This relation requires that what  $x$  believes is consistent not only with a certain proposition, but also with  $x$ 's own perspective towards such proposition.

## The Semantics of Attitudes De Se

Now that we have rephrased Hintikka's semantics in terms of centered propositions, we have the logical structure we need to finally account for the specific truth-conditions of de se propositional attitudes. If we apply our new analysis of "believe" to (3), repeated below, we obtain the truth-conditions in (5).

- (3) Kazimir believes that he himself is a painter  
 (5) (3) is true if, and only if,  $BEL_{Kazimir} \subseteq p_c$  (where  $p$  is the set of possible worlds where Kazimir belongs to the set of painters and  $c$  is Kazimir's own perspective)

According to the truth-conditions in (5), sentence (3) is true if, and only if, what Kazimir believes to be the case from his own perspective entails that Kazimir, as acknowledged from Kazimir's perspective, is a painter. To obtain truth-conditions that are strictly de se we must now add our final ingredient, *self-acquaintance*. As we know, what characterizes a de se belief is the fact that the believer is acquainted with the *res* the belief is about as herself. In the case of (3), for example, Kazimir must be acquainted with the individual he believes to be a painter as himself. All we need to do, then, to fully capture the truth-conditions of a de se belief, is to add to the truth-conditions in (5) the extra condition that the *res* the belief is about must be anchored to the center of the proposition by the relation of self-acquaintance. If we do so, we obtain the truth-conditions in (6).

- (6) (3) is true if, and only if, the following two conditions are met:  
 (i)  $BEL_{Kazimir} \subseteq p_c$   
 (ii) the holder of perspective  $c$  is acquainted with Kazimir as himself

(where  $p$  is the set of possible worlds where Kazimir belongs to the set of painters and  $c$  is Kazimir's own perspective)

According to (6), the truth of (3) requires that two conditions are met. First, what Kazimir believes to be the case from his own perspective must entail that Kazimir, as acknowledged from Kazimir's perspective, is a painter. Second, this perspective must be one of self-acquaintance. This second condition ensures that the belief is strictly *de se*. In fact, this is the condition that is contributed by control structures and logophoric pronouns whose role is, then, that of anchoring the *res* the belief is about to the center from which it is acknowledged.

In a sense, control structures and logophoric pronouns perform the same function as indexicals—they anchor a content to a context. However, whereas indexicals perform this function in the context of the actual circumstances of their use, control structures and logophoric pronouns perform their task within the local center of a proposition.

## Irrational Kazimir

By combining the notion of perspective with that of possible world, we are able to provide a richer notion of proposition and, with it, unambiguous truth-conditions for *de se* propositional attitudes. Context and acquaintance are not only factors that determine the meaning of linguistic expressions, as is the case with indexicals. They are also meaning themselves—things language is about. This is what we do, according to the framework we have developed, when we describe the specific first-personal perspective of someone's cognitive stance with a report of a *de se* propositional attitude.

The analysis of propositional attitudes we presented—based on centered possible worlds and centered propositions—was originally designed for *de se* propositional attitudes, but it has since been extended to other types of propositional attitudes. As demonstrated by a famous logical puzzle, originally due to Quine, all *de re* propositional attitudes involve, in fact, some relation of acquaintance between the subject of the attitude and the *res* the attitude is about. Hence, all *de re* propositional attitudes should ultimately be analyzed in terms of centered possible worlds and acquaintance relations.

Quine's puzzle is also known as the *double-vision problem*. Imagine a cognizing individual, say Kazimir, being presented to the picture in Fig. 26.4 and asked to judge what it depicts. Let us imagine that, as many of us would, Kazimir answers that Fig. 26.4 is the depiction of an elephant. Asked whether he believes the picture depicts some other animal, say, a swan, Kazimir confirms that, in his eyes, the picture portrays an elephant and not a swan. We can report the beliefs Kazimir forms after observing Fig. 26.4 with the following sentences:

- (7) Kazimir believes that Fig. 26.4 depicts an elephant
- (8) Kazimir believes that Fig. 26.4 does not depict a swan

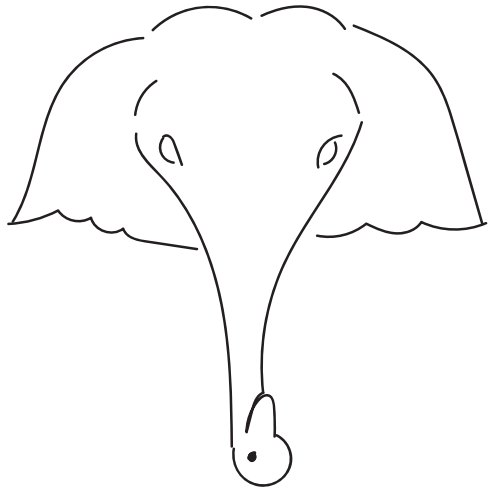
Both sentences are true of the belief state that Kazimir has come to endorse as a result of observing Fig. 26.4. Notably, both sentences are reports of a *de re* belief.

More precisely, they report two different *de re* beliefs Kazimir has come to endorse about the same *res*—Fig. 26.4.

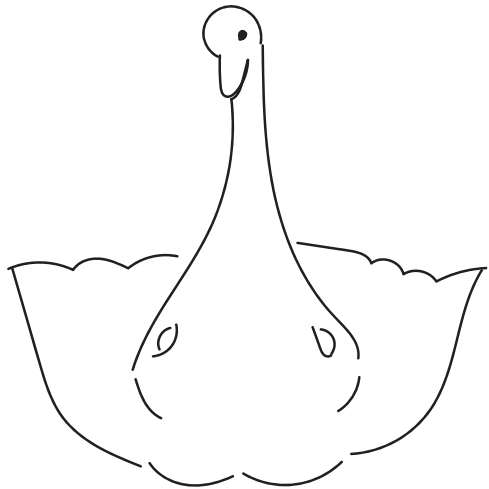
The scenario continues. After some time, Kazimir is presented to the picture in Fig. 26.5. Again, he is asked to judge what it depicts. Kazimir answers that Fig. 26.5 is the depiction of a swan. Asked whether he believes the picture depicts some other animal, say, an elephant, Kazimir confirms that, in his eyes, Fig. 26.5 portrays a swan and not an elephant. We can report the beliefs Kazimir has come to endorse about Fig. 26.5 with the following statements:

- (9) Kazimir believes that Fig. 26.5 depicts a swan
- (10) Kazimir believes that Fig. 26.5 does not depict an elephant

**Fig. 26.4** A picture of an elephant



**Fig. 26.5** A picture of a swan



As (7) and (8) were true in the case of Fig. 26.4, so (9) and (10) are now true for Fig. 26.5. Kazimir now believes that Fig. 26.5 is the depiction of a swan, but not that of an elephant. Notably, also (9) and (10) report a *de re* belief, although, this time, about Fig. 26.5.

There is, of course, a trick in the scenario. As the attentive reader may have already noticed, Figs. 26.4 and 26.5 are not two distinct pictures but, in fact, the same picture presented upside-down. Depending on its orientation, it can appear as the depiction of either an elephant or a swan (you can experiment with it by turning the book upside down and see the elephant become a swan, and vice versa). Because of the material identity between the two pictures, the scenario raises a problem for our current theory of *de re* propositional attitudes. More precisely, it forces us to the implausible conclusion that Kazimir is attributing contradictory properties to the same object and, therefore, is an irrational being—in the sense that he holds beliefs that are logically contradictory.

Let us see why this is so. Consider again sentence (7), repeated below. According to the analysis of *de re* propositional attitudes we have adopted so far, based on Hintikka's framework, *de re* propositional attitudes are relations between a subject and a *de re* proposition. Accordingly, (7) is true if, and only if, Kazimir's belief state is consistent with the proposition that the *res* corresponding to Fig. 26.4 satisfies the property of being the depiction of an elephant. However, given that Figs. 26.4 and 26.5 are one and the same object, the *de re* proposition that Fig. 26.4 is the depiction of an elephant is logically equivalent to the *de re* proposition that Fig. 26.5 is the depiction of an elephant. This means that (7) is logically equivalent to (11). According to our current understanding of *de re* propositional attitudes, in fact, the two sentences represent the same relation of propositional attitude between the same subject—Kazimir—and the same content—the proposition that the picture in question, call it Fig. 26.4 or Fig. 26.5, is the depiction of an elephant. Notice, however, that also sentence (10), repeated below, is true. But then, if sentence (10) is true and sentence (11) is also true, their conjunction, in (12), must be true as well.

- (7) Kazimir believes that Fig. 26.4 depicts an elephant
- (11) Kazimir believes that Fig. 26.5 depicts an elephant
- (10) Kazimir believes that Fig. 26.5 does not depict an elephant
- (12) Kazimir believes that Fig. 26.5 depicts an elephant and that Fig. 26.5 does not depict an elephant

According to (12), the object of Kazimir's belief is the proposition that one and the same object—Fig. 26.5—does and, at the same time, does not depict an elephant. As there are no logically consistent possible worlds where an object has and, at the same time, does not have the same property, this proposition corresponds to the *empty set*. We are ultimately led to the conclusion that Kazimir is an irrational creature holding contradictory beliefs.

This is obviously incorrect. We grant it, Kazimir has attributed contradictory properties to one and the same object. However, the reason why he did so is that he became acquainted with the same object in two different manners in the two differ-



ent circumstances in which he was presented to it. That is, Kazimir is holding two different beliefs about the same object *as contemplated from two different perspectives*.

## Acquaintance Generalized

Once again, the solution to the puzzle resides in the notion of perspective. To fully represent a de re belief of the type Kazimir has, we must consider not only the *res* the belief is about and the property Kazimir attributes to it, we must also consider the specific relation of acquaintance from which Kazimir acknowledges the *res* in question. The lesson we draw from Quine's double-vision problem is, therefore, that not only de se but, in fact, all de re propositional attitudes, are anchored to the center of the propositional attitude by a relation of acquaintance. De se attitudes are, after all, but one case of de re attitudes—the case where the relation of acquaintance at stake is one of self-acquaintance. The other de re attitudes differ only in that they are grounded on relations of acquaintance of different sorts.

Practically, this means that we can generalize our semantics of de se beliefs—based on centered propositions and acquaintance—to all de re beliefs:

- (13) A sentence of the form “ $x$  believes that  $y$  is  $P$ ” (where “ $y$  is  $P$ ” is a de re proposition attributing the property  $P$  to the *res*  $y$ ) is true if, and only if, the following two conditions are met:
- (i)  $BEL_x \subseteq p_c$
  - (ii)  $A(c, y)$
- (where  $p$  is the set of possible worlds where  $y$  is  $P$  and  $A$  is a relation of acquaintance)

According to these truth-conditions, a sentence of the form “ $x$  believes that  $y$  is  $P$ ” is true if, and only if, the following two conditions are met. First,  $BEL_x$ —the set of centered possible worlds that are consistent with what  $x$  believes from  $x$ 's perspective—must entail  $p_c$ —the set of centered possible worlds where  $y$  is  $P$  from perspective  $c$ . Second,  $y$ —the *res* the belief is about—must be acknowledged from  $c$  on the basis of a relation of acquaintance  $A$ . Different de re beliefs are characterized by different choices of  $A$ —the acquaintance relation. If the relation at hand is one of self-acquaintance, we have a de se belief. If it is a different relation of acquaintance we have a de re belief.

These truth-conditions allow us to explain why Kazimir is not an irrational creature. Whereas he is indeed attributing contradictory properties to the same object, he is nonetheless doing so while being acquainted with it in different ways. According to our new analysis, (7) and (11) are, in fact, no longer equivalent, as the two sentences relate Kazimir to two different centered propositions.

- (7) Kazimir believes that Fig. 26.4 depicts an elephant
- (11) Kazimir believes that Fig. 26.5 depicts an elephant

Sentence (7) attributes to Kazimir a belief in the centered proposition that the picture he is acquainted with as Fig. 26.4 is that of an elephant. Sentence (11) attributes to Kazimir a belief in the centered proposition that the picture he is acquainted with as Fig. 26.5—is that of an elephant. These are different centered propositions because, despite attributing the same property to the same object, they do so from different perspectives.

## References and Remarks

The notion of centered possible world was first contemplated by Quine (chapter 6 of Quine 1969) but was first adopted as a solution to the problem of *de se* propositional attitudes by Lewis (Lewis 1979).

The so-called double-vision problem was first illustrated by Quine in his article “Quantifiers and Propositional Attitudes” (Quine 1956). Quine’s original examples concerns an observer—Ralph—who becomes acquainted with the same individual—Orcutt—under two different sets of circumstances. On the one hand, Ralph has observed a man wearing a brown hat under questionable circumstances and has come to form the belief that this man is a spy. On the other hand, Ralph has become acquainted with a grey-haired man known as a pillar of the community and has come to form the belief that this man is not a spy. Unbeknownst to Ralph, the two men he observed are, in fact, the same individual—Orcutt. As in the example discussed in the chapter, if the content of Ralph belief is expressed in the form of a simple *de re* proposition, Ralph is necessarily attributed the belief in the contradictory proposition that Orcutt is and is not a spy.

As pointed out to us by an anonymous reviewer of this book, our example of the upside-down swan-elephant exploits a weakness of the human visual system, which excels at identifying right-side-up inversions but is also dismally incompetent at identifying upside-down inversions. This is especially clear in the domain of facial recognition. Interestingly, the same is observed in honeybees (see Dyer et al. 2005).

That acquaintance plays a role in *de re* reports was already observed by Russell in his “On Denoting” (Russell 1905), although it was Kaplan who first formulated explicit truth-conditions for *de re* propositional attitudes that include quantification over acquaintance relations (Kaplan 1968). The view of *de se* as a special case of *de re* is already contemplated by Lewis (Lewis 1979, p. 543) although the first generalized formal accounts of *de re* and *de se* propositional attitudes is offered by Cresswell and von Stechow in their article “*De re* belief generalized” (Cresswell and von Stechow 1982). In this chapter, we only discussed how centered worlds and acquaintance can contribute to a generalized definition of the truth-conditions of *de re* and *de se* propositional attitudes. We did not discuss how these truth-conditions can be produced compositionally. The literature offers a range of different proposals in this respect, often implemented in different frameworks but equally successful in enabling the notion of acquaintance to enter the compositional syntax (see, among others, Anand 2006; Chierchia 1989; Cresswell and von Stechow 1982; Maier 2009; von Stechow 2003; Schlenker 2004).

## References

- Anand, P. (2006). De se. PhD dissertation, MIT.
- Chierchia, G. (1989). Anaphora and attitudes de se. In R. Bartsch, J. van Benthem, & P. van Emde Boas (Eds.), *Semantics and contextual expressions* (pp. 1–31). Dordrecht: Foris.
- Cresswell, M. J., & von Stechow, A. (1982). De re belief generalized. *Linguistics and Philosophy*, 5(4), 503–535.
- Dyer, A. G., Neumeyer, C., & Chittka, L. (2005). Honeybee (*Apis mellifera*) vision can discriminate between and recognise images of human faces. *Journal of Experimental Biology*, 208, 4709–4714; <https://doi.org/10.1242/jeb.01929>
- Kaplan, D. (1968). Quantifying. *Synthese*, 19(1-2), 178–214.
- Lewis, D. (1979). Attitudes de dicto and de se. *Philosophical Review*, 88, 513–543.
- Maier, E. (2009). Presupposing acquaintance: A unified semantics for de dicto, de re and de se belief reports. *Linguistics and Philosophy*, 32(5), 429–474.
- Quine, W. V. O. (1956). Quantifiers and propositional attitudes. *The Journal of Philosophy*, 53, 177–187.
- Quine, W. V. O. (1969). *Ontological relativity and other essays*. New York: Columbia University Press.
- Russell, B. (1905). On denoting. *Mind*, 14(56), 479–493.
- Schlenker, P. (2004). A plea for monsters. *Linguistics and Philosophy*, 26, 29–120.
- von Stechow, A. (2003). Feature deletion under semantic binding. In M. Kadowaki & S. Kawahara (Eds.), *Proceedings of NELS 33* (pp. 133–157). Amherst, MA: GLSA Publications.

## Chapter 27

# Meaning and Epistemic Subjectivity



In the last chapters, we saw that perspective is a pervasive property of natural language meaning. We also saw that the notion of perspective we find in language—similarly to the notion of perspective we find in the visual arts—encompasses two dimensions. The first dimension concerns the causal relations that link meaning to the environmental conditions in which it is expressed. The second dimension concerns the psychological relation that connects meaning to the cognitive agents who express it. In this chapter, we will discuss what implications such notion of perspective has for the thesis of semantics externalism.

As we know, semantic externalism is the thesis that the prime ingredients of linguistic meaning are to be found in the external world of natural objects and not in the mind of the cognitive agents who entertain it. Is the notion of perspective in linguistic meaning compatible with such claims? As the notion of perspective is twofolded, so is the answer to this question. On the one hand, the causal link that binds meaning to its material environment can be regarded as a further argument for semantic externalism. The observation that meaning is inextricably bound to its context by a chain of causal relations proves the point of semantic externalism. It tells us that meaning is a function of the environment in which it occurs. On the other hand, the notion of acquaintance, being grounded in the psychology of the observer, brings with it an element of *subjectivity*. Acquaintance entails that language is not simply a means to describe the world. It also allows speakers to describe the world from the perspective of an observing cognitive agent.

Not surprisingly, then, acquaintance contributes a subjective dimension to linguistic meaning. Yet, we should be careful in judging the exact type of subjectivity that acquaintance involves. In Chap. 20, we distinguished between two types of subjectivity: *epistemic* and *ontological*. The first has to do with knowledge. We say that knowledge is subjective whenever it depends on the perspective of the knower. The second has to do with the nature of things. We say that things such as sensations, emotions, memories, and perceptual impressions are subjective because their very existence depends on that of the thinking mind who experiences them. Without

doubt, the notion of acquaintance introduces an element of epistemic subjectivity in meaning. If we are correct, natural language allows its speakers to communicate content as it is acknowledged from the vantage pointing of an observer. What is not as obvious is whether acquaintance also entails an element of ontological subjectivity. In favour of a positive answer, one may claim that the notion of acquaintance entails the existence of an observer endowed with the capacity to entertain subjective knowledge. What sort of object is an observer endowed with such capacity if not a conscious subject with a thinking mind? In reaction to these considerations, one may object that the step from epistemic subjectivity to ontological subjectivity is not so easily warranted. This is especially true in the light of Russell's definition of acquaintance. As we saw in Chap. 24, Russell took acquaintance to correspond to the *passive* registration of a sensorial input. According to this definition, acquaintance simply requires an observer that is endowed with the capacity to register information from the environment. It does not require that the observer is also endowed with a mind to reflect on the information collected. In fact, under Russell's definition, also a thermometer or a photo camera are endowed with the capacity to be acquainted with their environment. The information they register is, indeed, bound to their perspective—the time and place in which they find themselves—and, hence, epistemically subjective. Yet, neither the thermometer nor the camera are endowed with a conscious mind. Hence, at least in its strict Russellian understanding, acquaintance does not automatically entail ontological subjectivity. The model that interprets natural language has the logical structure to articulate the subjective perspective of an observer but this does not mean that the model is populated by objects that are themselves of a subjective nature.

The safest and most modest conclusion to be drawn at this point is, therefore, that acquaintance entail epistemic subjectivity but not necessarily ontological subjectivity. Indeed, some readers may find the Russellian notion of acquaintance, defined as a strictly mechanical relation between an observer and its environment, unintuitive and unconvincing. They should know that they are not alone. To be sure, there are major reasons for concerns with it, beginning with the fact that the sharp distinction Russell draws between the *passive* activity of capturing information through the senses and the *active* process of cognizing about it is at odds with what we know today about the physiology and psychology of perception. We will return to this objection in part III where we will discuss at length how the recent advances in the cognitive sciences have forced scientists to abandon the more traditional views of perception, such as the Russellian one. These findings, as we will see, provide a very different view of perception and, in particular, of the role perception plays at the interface between world and mind.

For the time being, however, we will focus on a different sort of problem with Russell's notion of acquaintance. In the coming chapters, we will consider a special case of *de se* propositions—whose discovery we owe, once again, to Wittgenstein—that cannot be accounted for in the material terms of Russell's definition of acquaintance. This special type of *de se* demonstrates that the “special and primitive”

process of identification that, according to Frege, allows speakers to refer to themselves in the first-person does not exhaust itself with the speakers' acquaintance with their environment, but also entails an element of *introspection*—and introspection does entail an element of ontological subjectivity.

## Chapter 28

# Implicit De Se



Our inquiry into the semantics of the first-person is not over yet. So far, observing how speakers use natural language to refer to themselves has enabled us to define the notion of perspective in language and, in particular, how sensorial acquaintance is a fundamental element of such notion. In this chapter, we will review a special type of *de se*, which entails a form of self-identification that is of an introspective nature and, therefore, cannot be captured by the notion of sensorial acquaintance. This special form of *de se* is commonly referred to as *implicit de se* and it was first observed by Wittgenstein.

### Error Through Misidentification

Let us begin by observing that the process of acquaintance, being grounded on the sensorial skills of the observer, brings with it the possibility of error.

Imagine that Frida sees a man from a distance and, after observing him more carefully, comes to the conclusion that the man is Kazimir and that he is wearing red shoes. She expresses her belief by uttering (1).

(1) Kazimir is wearing red shoes

The declarative sentence in (1) expresses the proposition that the individual Kazimir belongs to the set of individuals who are wearing red shoes. It is a *de re* proposition, attributing a property—that of wearing red shoes—to a *res*—Kazimir.

Because the proposition expressed by (1) is constituted of these two elements—a *res* and a property—there are two ways in which Frida may be mistaken in her claim. One possibility is that Frida is attributing an incorrect property to Kazimir. Suppose, for example, that Frida is correct in thinking that the person she saw is Kazimir, but, having observed him from a distance, mistook what he is wearing for a pair of red shoes when they are, in fact, boots. In such scenario, Frida would be

correct in identifying Kazimir as the *res* her belief is about but incorrect in attributing to him the property of wearing red shoes. The second possibility is that the proposition attributes the right property to the wrong *res*. Suppose that the man Frida is looking at is indeed wearing red shoes but, contrary to what Frida thinks, he is not Kazimir, but Pablo. In such case, Frida would be correct in attributing the property of wearing red shoes to the man she sees but wrong in identifying that man as Kazimir.

Of course, the two possible errors are not mutually exclusive and can occur together. It could very well be that Frida is mistaken both in identifying the *res* her belief is about as Kazimir and in attributing to him the property of wearing red shoes. In this chapter, however, our main concern will be the second sort of error—that which is based on the misidentification of a *res*. We will refer to this type of error as an *error through misidentification*. It is the type of error we make when we form a correct judgment that there is a *res* that satisfies a certain property, but we make a mistake in identifying who (or what) that *res* is.

In principle, all de re propositions are susceptible to error through misidentification. In fact, errors through misidentification can also occur with de se propositions—that is, de re propositions one entertains about oneself with awareness of doing so. Although only under special circumstances, it is indeed possible to make mistakes even when identifying ourselves as such. As an example, consider sentence (2), which attributes the property of wearing red shoes to the individual the speaker is acquainted with as herself.

(2) I am wearing red shoes

It is not difficult to think of a scenario where the speaker's de se judgment is wrong because of an error through misidentification. Suppose that the sentence is uttered by Kazimir when observing an old picture of a little boy wearing red shoes and believing, erroneously, that the boy in the picture is he himself when he was a boy. In this scenario, Kazimir's utterance of (2) is false because, although Kazimir correctly characterized the individual in the picture as wearing red shoes, he was mistaken in acknowledging that that individual is he himself and, therefore, in referring to him in the first-person.

The observation that natural language lends itself to errors through misidentification is not surprising. It is, in fact, what we should expect given the framework we have developed in the previous chapters. As we saw, speakers are connected to the *res* their propositions are about by a relation of *acquaintance*. Acquaintance, however, is a form of sensorial perception and, as our senses can be deceived into error, so acquaintance can be an erroneous source of information. Erroneous acquaintance with her environment is what deceives Frida into believing that the man she sees is Kazimir, when, in fact, it is someone else. It is also what deceives Kazimir into identifying the boy he sees in the picture as he himself, when it is somebody else. It is what deceived the ancient Greeks into believing that the last star they saw in the morning and the first star they saw in the evening are two different objects, when, in fact, they are one and the same. As acquaintance is susceptible to error, so are the expressions of language whose meaning is grounded in it.



## Immunity to Error Through Misidentification

Against these seemingly obvious considerations, Wittgenstein observed that some de se statements are *immune* to error through misidentification—that is, speakers cannot be mistaken in identifying themselves as the *res* their statements are about. Some relevant examples are provided by the sentences in (3)–(5):

- (3) I feel cold
- (4) I am happy
- (5) I see a black square

These sentences are all in the first-person. In the terms of the analysis we have developed so far, they all express de se propositions—that is, propositions about a *res* that the speaker identifies with herself.

Wittgenstein observes that these sentences differ from the de se sentences we have considered so far in that they are not susceptible to error through misidentification. To prove the point, Wittgenstein asks us to observe the following difference between the sentences in (3)–(5) and those in (1)–(2). Suppose that Frida declares to us, while seeing someone from a distance, “Kazimir is wearing red shoes”, but we doubt the person Frida sees from a distance is Kazimir. We can always react by saying, “yes, somebody is certainly wearing red shoes, but are you sure it’s Kazimir?” Similarly, if Kazimir says, “I am wearing red shoes”, while looking at a picture, but we doubt the individual he is talking about is Kazimir himself, we can reply, “surely the boy in the picture is wearing red shoes, but are you sure it’s you?” We can question whether the speaker is correct in identifying the *res* her utterance is about precisely because, as we saw above, these sentences are susceptible to error through misidentification. The same observation, however, does not apply to the sentences in (3)–(5). It would be odd, to say the least, to challenge someone who says, “I feel cold”, by replying, “surely somebody is feeling cold, but are you sure it’s you who is feeling cold?” It would be equally odd to react to someone saying, “I am happy”, by replying, “Surely someone is happy, but are you sure it’s really you who is happy?” And the same goes for “I see a black square”. The sentences in (3)–(5) appear to be *immune to error through misidentification* in the sense that they cannot be false because the speaker mistakenly identified the individual feeling cold, happy, or seeing a black square as herself when, in fact, it is someone else.

In this chapter, we will see that the immunity to error displayed by these sentences has to do with the fact that they express a peculiar type of de se, which, following the terminology of the philosopher François Recanati, we will refer to from now on as *implicit* de se.

## Introspection

The discovery that some de se propositions are immune to error through misidentification is extremely important because it demonstrates that acquaintance is not the sole mode of identification of the referent of a linguistic expression.

There is, of course, an intuitive explanation of why sentences such as (3)–(5) are immune to error through misidentification: The type of identification they involve is not based on perceptual acquaintance but on mental *introspection*. In favor of this type of explanation, notice that the sentences in (3)–(5) are all characterized by two common properties: They are all in the first-person and their predicates—“feel cold”, “be happy”, and “see a black square”—all refer to *psychological states*—sensations, emotions, perceptual experience. Both these properties are necessary conditions for immunity to error through misidentification to arise. If we were to replace the first-person subjects of the sentences in (3)–(5) with nominal expressions in the second or third-person, we would obtain sentences that are not immune to error through misidentification. Consider, as an example, sentence (6), which is obtained from sentence (3) by replacing the first-person subject “I” with the third-person subject “Kazimir” (and, of course, by adapting the verbal morphology accordingly).

(6) Kazimir feels cold

Sentence (6) is not immune to error through misidentification. It would be perfectly legitimate, in the appropriate circumstances, to challenge someone claiming, “Kazimir feels cold”, by replying, “surely someone is feeling cold, but are you sure it’s Kazimir?” Indeed, when attributing the property of feeling cold to someone other than ourselves, we can make mistakes due to erroneous identification. Similarly, if we were to replace the predicates of psychological state in (3)–(5) with predicates that are not of psychological state, we would also obtain sentences that, despite being in the first-person, are not immune to error through misidentification. We already provided such an example in (2), repeated below.

(2) I am wearing red shoes

Even though sentence (2) is in the first-person, it is not immune to error through misidentification. As we saw above, there are perfectly legitimate circumstances in which it is possible to challenge someone claiming, “I am wearing red shoes”, by replying, “surely someone is wearing red shoes but are you sure it’s you?”

Together, these two properties support the view that the reason behind the special status of sentences such (3)–(5) is that they express self-reflective judgements made by the speaker about her own inner psychological life. Typically, people do not identify themselves as the subjects of their own sensations, emotions, and perceptual experience on the basis of sensorial acquaintance. They rather do so by becoming introspectively aware of their inner psychological lives. We do not realize that we are feeling cold, happy, or perceiving a black square by observing our external appearance or behavior. We rather do so by becoming aware of what goes on in our minds.

Differently from sensorial acquaintance, introspection is immune to error through misidentification. Whereas we can make mistakes in identifying objects through our senses, we can never fail in identifying ourselves as the subjects of our own sensations, emotions, and perceptual experience. When we feel cold, we cannot fail in recognizing ourselves as the ones experiencing such sensation. We may be wrong in classifying our experience as one of feeling cold, but we cannot be

wrong in believing that, whatever we happen to be feeling, it is us who are feeling it, and not somebody else. This is why sentences such as (3)–(5) are immune to error through misidentification.

The idea that implicit de se statements are immune to error through misidentification because they are based on identification through introspection is simple and intuitive. Yet, as the attentive reader may have already noticed, it comes with serious consequences for semantic externalism. Introducing the notion of introspection in the model of interpretation of natural language also means introducing an element of ontological subjectivity and, thereof, privacy. It is equivalent, in effect, to admitting that the sentences in (3)–(5) are exemplars of a private language. We will discuss these implications in the following chapter. For the time being, we shall observe that the phenomenon of the implicit de se is pervasive in natural language.

## Implicit De Se Propositional Attitudes

In Chap. 25, we saw that, as there are de se propositions, so there are de se propositional attitudes. We will now see that, as there are implicit de se propositions, so there are implicit de se propositional attitudes. In English, we find unambiguous expressions of implicit de se attitudes with verbs such as “imagine”, “dream”, and “remember”. As an illustrative example, we will shall consider the verb “remember”. Consider first sentence (7).

(7) Kazimir remembers that he gave the speech on the future of art

In (7), the verb “remember” is used like a regular verb of propositional attitude, such as “believe”. It combines with a subject—“Kazimir”—and an inflected complement clause—“he gave the speech on the future of art”—to express a relation of propositional attitude between an individual and a centered proposition. As we have noticed in the case of other verbs of propositional attitude, sentence (7) is ambiguous in two ways. Firstly, the pronoun “he” can take different referents in different contexts, hence, the sentence can express different de re propositional attitudes in different circumstances. Secondly, when “he” is intended to refer to the subject—Kazimir—the sentence remains ambiguous between a de re and a de se interpretation. The sentence tells us that Kazimir has a memory of Kazimir giving the speech, but does not tell us whether Kazimir is aware of the identify between himself and the *res* his memory is about. There are two ways to unambiguously express a de se propositional attitude with the verb “remember”. The first is (8), which is by and large equivalent to (7), except in that it features the *logophoric* pronoun “he himself” in place of the regular personal pronoun “he”.

(8) Kazimir remembers that he himself gave the speech on the future of art

The second is the *control* structure in (9), which differs from (7) in that its embedded clause “giving the speech on the future of art” features an uninflected verb in the gerundive and no explicit subject.

(9) Kazimir remembers giving the speech on the future of art

Both (8) and (9) unambiguously express a de se propositional attitude. They both describe a memory that Kazimir holds about himself in full awareness of doing so. Yet, they are not completely equivalent. Jerry Fodor was the first to observe that control structures like (9) have peculiar logical properties. Consider the inferential schema in (10). It reports a logical inference that derives a conclusion *C* on the basis of two premises: *P*<sup>1</sup>, which corresponds to sentence (9), and *P*<sup>2</sup>.

(10) *P*<sup>1</sup> Kazimir remembers giving the speech on the future of art

*P*<sup>2</sup> Only Kazimir gave the speech on the future of art

---

*C* Only Kazimir remembers giving the speech on the future of art

The argument is valid. No one other than Kazimir could possibly remember giving the speech, because no one other than Kazimir gave the speech. No one could possibly reply to the conclusion by saying “Hey, I also remember giving the speech on the future of art”, because doing so would contradict the second premise. Interestingly, the same logical inference fails if we use (8) instead of (9):

(11) *P*<sup>1</sup> Kazimir remembers that he himself gave the speech on the future of art

*P*<sup>2</sup> Only Kazimir gave the speech on the future of art

---

*C* Only Kazimir remembers that he himself gave the speech on the future of art

In (11), we find the same inferential schema of (10). The only difference is that we have now replaced all the occurrences of sentence (9) with sentence (8). This time, however, the inference is not valid. Even if it is indeed the case that “Kazimir remembers that he himself gave the speech on the future of art” and Kazimir is the only one who gave the speech in such occasion, it does not follow, as a matter of logical deduction, that “only Kazimir remembers that he himself gave the speech on the future of art”. This is because other people—for example, those that witnessed the speech—may very well remember that Kazimir himself gave the speech. The conclusion does not follow from the premises.

The comparison between the two arguments demonstrates that the two sentences (8) and (9) are not logically equivalent. Further proof of this comes from the observation that (8) can be conjoined with the negation of (9) without producing a contradiction.

(12) Kazimir remembers that he himself gave the speech but he does not remember giving the speech

Intuitively, we can describe the difference in meaning between the two sentences as follows. Sentence (8) reports a memory that Kazimir has of himself from a perspective that other observers can share. As Kazimir remembers that Kazimir—whom Kazimir acknowledges as he himself—gave the speech on the future of art, so others can remember that Kazimir gave the speech on the future of art. Conversely,

sentence (9) reports Kazimir’s memory from a subjective perspective, which only Kazimir can entertain and no one else can share with him. Except for Kazimir himself, no one else can remember Kazimir’s subjective experience of giving the speech.

Once again, what seems special about the meaning of (9) is the peculiar subjective perspective from which it reports Kazimir’s memory. Also in this case, it is intuitive to identify such perspective with introspection. Sentence (9), in effects, reports a memory that Kazimir has of himself that is based on his own introspective experience of giving the speech. It tells us that Kazimir remembers the actual experience of giving the speech “from within”—an expression we borrow from the philosopher James Pryor. In our terms, sentence (9) expresses an implicit *de se* propositional attitude—a memory Kazimir has of his introspective experience of giving the speech.

## Evidentiality

Another domain of language where we find evidence that the implicit *de se* is linguistically relevant is that of *evidentiality*.

If we were to translate sentence (3)—“I feel cold”—in Japanese, we will obtain sentence (13), where “*watashi*” means “I”, “*samui*” means “being cold”, and “*wa*” is a marker of topicalization (roughly) indicating that “*watashi*” is what the sentence is about.

- (13) *Watashi wa samui*  
“I am cold”

Sentence (13) enjoys a very peculiar property. It is the only grammatical structure of Japanese where the predicate “*samui*” can occur freely. If “*samui*” is combined with a subject that is not the first-person pronoun “*watashi*”, it produces an ungrammatical sentence. If combined, for example, with third-person pronoun “*kare*” (English “he”)—as in (14)—it is judged by speakers of Japanese as ungrammatical.

- (14) *Kare wa samui*  
“He is cold”

The correct way to say “he is cold” in Japanese is not (14) but (15), where “*samui*” is accompanied by the expression “*noda*”.

- (15) *Kare wa samui noda*  
“He is cold”

“*Noda*” is what linguists call an *evidential*. Evidentials are a grammatical category of linguistic expressions—attested in a number of languages of the world—whose role is to express the source of the evidence a speaker has when asserting a

content. For example, evidentials can tell us whether the content asserted has been directly witnessed—seen or heard—or only inferred from indirect evidence. The Japanese “Noda”, to return to our case in point, marks the fact that the speaker has only indirect evidence for the proposition she is asserting, where “indirect” means that her assertion is not supported by direct observation but the result of a process of inference. Literally, (15) means something such as, “*as far as I can tell*, he is cold”.

Why is the evidential “noda” obligatory in the case of (15) but not in the case of (13)? That is, why must Japanese speakers use “noda” whenever combining a predicate such as “samui” with a subject that is not in the first-person but omit it when combining the same predicate with a first-person subject? There is an intuitive explanation of these facts. “Samui” is a predicate of psychological states, which attributes to its subject the property of being the subject of an experience of feeling cold. As a psychological state can only be accessed through introspection, a statement that somebody feels cold can only be made in the first-person. If we wish to make a statement that somebody else is feeling cold, Japanese grammar enforces us to declare, by use of the evidential “noda”, that the judgment is based on indirect evidence. The reason why “samui” can occur without the evidential only with the first-person is, therefore, that only in such case it expresses an implicit de se proposition based on introspection. When used with a different person, it expresses a judgment that is based on indirect evidence and, therefore, susceptible to error. This is a fact that must be explicitly marked in Japanese.

Evidentials, as they are attested in Japanese and other languages, offer yet another argument that natural language is capable of expressing the introspective knowledge a subject has of her inner psychological life. Of course, these findings pose a significant challenge to semantic externalism. We turn to this problem in the following chapter.

## References and Remarks

Wittgenstein’s observations on immunity to error through misidentification were first made in in “The Blue and Brown Books”, two collection of notes that Wittgenstein made between 1933 and 1935 for private and pedagogic use. These notes circulated informally during Wittgenstein’s time but were never published. They were published posthumously in 1958 with the title *Preliminary Studies for the “Philosophical Investigations”, Generally known as The Blue and Brown Books* (Wittgenstein 1958). The section that is relevant to our discussion follows page 66 of the original page division. In his notes, Wittgenstein does not explicitly use the term “immunity to error through misidentification” but distinguishes between two uses of the first-person—I as object and I as subject—the former characterized by the possibility of error of reference, the second immune to it. The label “immunity to error through misidentification” was introduced by Sydney Shoemaker in his 1968 article “Self-reference and self-awareness” (Shoemaker 1968). Shoemaker’s article was pivotal to introducing the topic to the forum of philosophical discussion.

Since Wittgenstein's observations and Shoemaker's essay, the topic of immunity to error through misidentification has witnessed a significant growth in attention from philosophers and, more recently, linguists. We refer the interested reader to the two collections *Self Reference and Self Awareness* (Brook and DeVidi 2001), which contains a number of classic essays on the topic including Shoemaker's homonymous 1968 article, and *Immunity to Error through Misidentification: New Essays* (Prosser and Recanati 2012), which collects a number of recent theoretical contributions on the topic from a variety of perspectives. The terminology "implicit de se" was introduced by François Recanati in his book *Perspectival Thought* (Recanati 2007). The qualification of implicit de se statements as expressing a content "from within" is taken from James Pryor's 1999 essay "Immunity to error through misidentification" (Pryor 1999).

Jerry Fodor's observations on the infinitival complement of "remember" are from Jerry Fodor's book *The Language of Thought* (Fodor 1975). His observations are discussed and extended in Chierchia (1989) and Higginbotham (2003, 2010). In particular, we owe to James Higginbotham the observation that the semantic properties of the gerundive and infinitival complements of verbs such as "remember", "imagine", and "want" are best explained in terms of the notion of immunity to error through misidentification. Higginbotham also offers an account of the implicit de se that is based on the framework of event semantics (for a brief introduction to event semantics, see the reference and remarks section of Chap. 14). According to Higginbotham's account, the reference of the silent subject of the embedded gerundive or infinitival—what is commonly referred to in the syntactic literature as PRO—is constrained by two conditions, one linking it to the event expressed by the embedded predicate, the other linking it to the event expressed by the matrix predicate. Exemplified on the sentence "Kazimir remembers giving the speech on the future of art", Higginbotham's approach contends that the silent subject of the embedded sentence "giving the speech on the future of art" is identified as the object that satisfies the following two conditions: It is the individual satisfying the thematic role assigned to it by the embedded predicate—that is, it is the agent in the event of giving the speech on the future of art—and is the subjective experiencer of the event expressed by the matrix verb "remember". The two conditions capture two central aspects of the meaning of the controlled subject. The first is that its reference is unambiguously identified as the same as the subject of the main verb. The second is that its reference is immune to error through misidentification. In our example, if Kazimir identifies the agent of the event of giving the speech with the subject of the experience of remembering that event, he cannot fail to identify this subject as he himself. Higginbotham's analysis offers a powerful framework for the analysis of the implicit de se. As we will see in the following chapter, however, his framework introduces an element of irreducible ontological subjectivity—which is incompatible with the tenants of semantic externalism. For a detailed presentation and critical assessment of Higginbotham's analysis, we refer the reader to Fiorin and Delfitto (2014). An attempt at framing Higginbotham's original observations and theoretical solutions into a comprehensive theory of control is made in Delfitto and Fiorin (2018).

The observations concerning Japanese evidentials are reported in Kuroda (1973), Kuno (1973), and Aoki (1986) and are further discussed in Tenny (2006).



## References

- Aoki, H. (1986). Evidentials in Japanese. In W. Chafe & J. Nichols (Eds.), *Evidentiality: The linguistic coding of epistemology* (pp. 159–167). Norwood, NJ: Ablex Publishing.
- Brook, A., & DeVidi, R. C. (2001). *Self reference and self awareness*. Amsterdam: John Benjamins Publishing.
- Chierchia, G. (1989). Anaphora and attitudes de se. In R. Bartsch, J. van Benthem, & P. van Emde Boas (Eds.), *Semantics and contextual expressions* (pp. 1–31). Dordrecht: Foris.
- Delfitto, D., & Fiorin, G. (2018). Where syntax meets pragmatics: Varieties of de se and control structures. *Intercultural Pragmatics*, 15, 161–190.
- Fiorin, G., & Delfitto, D. (2014). Control, attitudes de se and immunity to error through misidentification. *Rivista Internazionale di Filosofia e Psicologia*, 5, 184–206.
- Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.
- Higginbotham, J. (2003). Remembering, imagining, and the first person. In A. Barber (Ed.), *Epistemology of language* (pp. 496–533). Oxford: Oxford University Press.
- Higginbotham, J. (2010). On words and thoughts about oneself. In F. Recanati, I. Stojanovic, & N. Villanueva (Eds.), *Context-dependence, perspective, and relativity* (pp. 209–238). Berlin: Mouton-De Gruyter.
- Kuno, S. (1973). *The structure of the Japanese language*. Cambridge, MA: MIT Press.
- Kuroda, S.-Y. (1973). Where epistemology, style, and grammar meet: A case study from Japanese. In S. Anderson & P. Kiparsky (Eds.), *A festschrift for Morris Halle* (pp. 377–391). New York: Holt, Rinehart and Winston.
- Prosser, S., & Recanati, F. (2012). *Immunity to error through misidentification: New essays*. Cambridge: Cambridge University Press.
- Pryor, J. (1999). Immunity to error through misidentification. *Philosophical Topics*, 26(1/2), 271–304.
- Recanati, F. (2007). *Perspectival thought: A plea for (moderate) relativism*. Oxford: Oxford University Press.
- Shoemaker, S. (1968). Self-reference and self-awareness. *Journal of Philosophy*, 65, 555–567.
- Tenny, C. L. (2006). Evidentiality, experiencers, and the syntax of sentience in Japanese. *Journal of East Asian Linguistics*, 15, 245–288.
- Wittgenstein, L. (1958). *Preliminary studies for the “philosophical investigations”, generally known as the blue and brown books*. London: Basic Blackwell.



## Chapter 29

# Meaning and Ontological Subjectivity



### Introspection and Ontological Subjectivity

The facts we reviewed in the previous chapter suggest that acquaintance is not the only type of epistemic relation speakers entertain with the content of their words. The phenomenon of immunity to error through misidentification—as we find it in implicit *de se* sentences and reports of propositional attitudes—suggests that speakers can identify themselves as objects of reference also on the basis of *introspection*. But what is introspection?

In a nutshell, introspection is the process that allows cognitive subjects to become aware of their inner psychological lives. It differs from acquaintance in significant ways. Acquaintance, as we know, relates subjects to their environment, the external reality that surrounds them. Amongst other things, it allows subjects to individuate objects in their environment, including themselves—in which case we talk about self-acquaintance. However, as acquaintance is the product of a sensorial apparatus—a mechanism capable of reacting causally to its environment—it is also susceptible to error. Introspection, conversely, relates subjects to their own psychological lives. It is private and reflexive, as it only allows subjects to individuate their own psychological states and not those that belong to others. It is an inner *mental* capacity and, as such, it is immune to error through misidentification.

In Chap. 27, we saw that the notion of acquaintance brings with it an element of *epistemic subjectivity*. Acquaintance means that the model described by natural language is anchored to a perspective. In the same chapter, we also observed that acquaintance does not necessarily entail an element of *ontological subjectivity*. That is, acquaintance does not mean that the model comprises objects of a subjective nature. This is especially true if we maintain—as Russell does—that acquaintance merely corresponds to the capacity to register environmental information through the senses. As such, acquaintance does not require that we introduce in our model an observer endowed with a sentient mind. The same, however, is not true of

introspection. To be capable of introspection, an individual must be endowed with a sentient mind. Hence, admitting that natural language has the resources to express judgments based on introspection is equivalent to admitting that the model of interpretation of natural language includes, amongst its essential ingredients, objects and properties whose very existence depends on that of a sentient mind and are, therefore, ontologically subjective.

This conclusion represents a formidable objection against semantic externalism as it is obviously incompatible with the claim that meaning is external to the minds of speakers. Admitting that implicit *de se* propositions are judgments based on introspection is equivalent to admitting that natural language has the capacity to refer to objects and properties that belong to the private psychological lives of speakers. But admitting this is equivalent to admitting that private language is possible and, therefore, that Wittgenstein's argument against private language is wrong.

## Reference and Error

According to some scholars, we should be careful in using Wittgenstein's observations on immunity to error through misidentification as an argument against semantic externalism. Some readers may have found already suspicious that the phenomenon of immunity to error through misidentification—which, so far, we have interpreted as evidence that private language is possible—was originally observed by Wittgenstein himself—the same thinker who provided the argument against private language. Did he not notice the incongruity between the two? According to the philosopher Elizabeth Anscombe—a major scholar and translator of Wittgenstein—it is a mistake to interpret the phenomenon of immunity to error through misidentification as evidence that the referent of the first-person pronoun is sometimes identified through introspection. Rather, it should be understood as evidence that the meaning of the first-person is not its reference at all.

The essential hypothesis at the basis of the theory of meaning as reference is that the meaning of a linguistic expression is the object it refers to. As we know, the relation that an expression holds to its referent is arbitrary and, therefore, contingent. That “Kazimir Malevich” means Kazimir Malevich is a matter of stipulation, not principle. However, the fact that the relation of reference is arbitrary and contingent entails that it is also *susceptible to error*, and necessarily so. As there is no conceivable referring expression whose association with its referent holds as a matter of necessity, there is always the possibility of making a mistake when using it. If this is correct, the lesson we should draw from the observation that some expressions are immune to error through misidentification is not that these expressions refer to private objects but, rather, that they do not refer at all. If meaning is reference, it must be susceptible to error. Hence, if we find an expression that is immune to error through misidentification, the conclusion we should draw is that its meaning is not its reference.

## Introspection and Identification

In favour of this interpretation, Anscombe points out a fundamental problem with the notion of introspection. An indispensable assumption we must make, if we wish to explain immunity to error through misidentification on basis of the notion of introspection, is that introspection is a relation of *identification*. Introspection, that is, must allow sentient beings not only to explore their inner psychological lives but also to identify themselves as the owners and experiencers of such psychological lives. In effect, the view we have entertained in the previous chapter is that a sentence such as (1) is immune to error through misidentification because the referent of “I”—that is, the object “I” refers to—is identified through a process of mental introspection.

(1) I feel cold

It is through the introspective experience of feeling cold that the speaker comes to identify herself as the object feeling cold. And it is through the same process that she comes to identify herself as the referent of the first-person pronoun “I”.

Against this view of introspection, stands one of the most notable arguments in the Western philosophical tradition: Descartes’ *skeptical argument*. In a famous passage of his “Meditations on First Philosophy”, the philosopher René Descartes invites us to consider the possibility that our experience of the world—the way we comprehend it through our minds—is a mere illusion. To make his point, Descartes invites us to contemplate the possibility that an evil demon is currently deceiving our minds and senses. Whatever it is that we are now experiencing—say, reading a book while lying comfortably on our favourite armchair or enjoying the beautiful colours of the trees while having a stroll in the park—it is not really happening. It is but a mere psychological illusion induced by the evil demon. This scenario could be easily dismissed as the outlandish fantasy of a philosopher, were it not for a quite extraordinary reason: If it was true, we could not tell. If, in effect, the world we now perceive and experience was a mere illusion induced by a deceptive demon, we would not be able to prove that it is so. The challenge provided by Descartes’ skeptical argument is, ultimately, an epistemic one. It shows us that, in and by itself, the introspective awareness we have of our inner psychological lives—crucially including the way we consciously experience the products of our senses—provides us with no reliable knowledge of what the world out there really is like. It could all be a mere illusion and, if it were, we could not prove it to be such.

Amongst its many implications, the skeptical argument also entails that introspection cannot be taken as a source of identification. Introspection certainly provides us with a sense of ownership of our own psychological lives. We acknowledge our feelings, sensations, and perceptual experience as belonging to us, not others. Yet, knowledge of our inner psychological lives, in and by itself, grants us no information about the sorts of objects we are in the real world out there.

To illustrate the relevance of Descartes' skeptical argument to the problem of self-identification, Anscombe offers a version of her own of the skeptical argument—known as the *tank argument*. Anscombe invites us to consider a thinking human being in a state of total sensory deprivation. Let us call this being Sensory Deprived Kazimir. For medical reasons, Sensory Deprived Kazimir is kept floating in a tank full of warm liquid that keeps him alive. In the tank, his senses are completely anaesthetized. He cannot see, hear, or feel anything. He is alienated from any information about who he is, where he is, what happened to him, or what he is doing there. Yet, as his brain is maintained alive and active, he can still think. In these extreme conditions, Sensory Deprived Kazimir is capable of introspection—that is, he is able to think as well as recognize his thoughts as his own. However, in and by itself, this capacity does not grant him any knowledge of who he is as a material object in a material world.

Anscombe relies on this thought-experiment to illustrate the point that introspection is strictly mental. In and by itself, it contributes no knowledge of the material world and, henceforth, does not allow introspective subjects to identify themselves as objects of the material world. The conclusion is that introspection cannot be taken to support, in and by itself, any form of self-identification. Being aware of our own thoughts tell us nothing of what sort of material objects we are.

Needless to say, these considerations have important philosophical implications. From our linguistic perspective, the most relevant implication is that we cannot say that the implicit *de se* is immune to error through misidentification because it is based on identification through introspection, for, if Descartes and Anscombe are right, introspection alone does not provide any identification at all.

## “I” as a Non-referring Term

The most important conclusion Anscombe draws from her observations is that the meaning of the first-person must not be identified with its reference. On the one hand, if its meaning were its reference, it should not be immune to error through misidentification. Reference, as we saw, entails the possibility of error. Hence, the lack of error must be understood as an indication of lack of reference. On the other hand, the introspective process that supports implicit *de se* judgments does not support, in and by itself, any form of identification. Again, this suggests that immunity to error through misidentification is due to the fact that the implicit *de se* involves no identification at all, rather than identification through introspection.

The idea that the first-person pronoun is a non-referring expression is intriguing. It certainly squares with Wittgenstein's more general plan, which we illustrated in previous chapters, to eliminate the very notion that language is a system of reference. Yet, it also meets some problems. To begin with, we should not forget that not all uses of the first-person are immune to error through misidentification. Should we then conclude that not all uses of the first-person are non-referential? Compare (1), repeated below, to (2).

- (1) I feel cold
- (2) I am wearing red shoes

As we saw in the previous chapter, only (1) is immune to error through misidentification. Sentence (2) is not. As it is a judgment based on acquaintance, the speaker may very well be mistaken in judging herself as the individual wearing red shoes. Anscombe’s suggestion that “I” is a non-referring expression explains why (1) is immune to error through misidentification, but what are we to do with (2)? How are we to account for the fact that (2) is not immune to error through misidentification?

The problem with Anscombe’s view is that, to explain the cases in which the first-person is immune to error through misidentification, it ends up neglecting the cases in which the first-person is *not* immune to error through misidentification. A possible solution to this problem requires maintaining that “I” is systematically ambiguous between a non-referring and a referring use. When not referring, it brings about a judgment that is immune to error through identification. When referring, it brings about a judgment that is not immune to error through misidentification. This solution, however, is undeniably stipulative. No attested natural language, as far as we know, displays a visible difference between the first-person as it is used in judgments that are immune to error through misidentification and the first-person as it is used in judgments that are not immune to error through misidentification. In all languages we know of, “I” as used in sentences such as (1) and “I” as used in sentences such as (2) are one and the same word. This conspires against the view that the first-person is systematically ambiguous between two such different interpretations.

A further problem for Anscombe’s proposal is that it explains immunity to error through misidentification by focusing exclusively on the first-person and its meaning. This, however, is not enough. As we saw in the previous chapter, immunity to error through misidentification is the product of two complementary factors—a first-person subject and a predicate of psychological state. Both elements are necessary conditions for immunity to error through misidentification to arise. Sentence (1) is immune to error through misidentification not only because its subject is in the first-person but also because its predicate—“feel cold”—expresses a property of psychological states. A satisfactory account of the meaning of (1) should tell us how these two factors contribute together to the fact that (1) is immune to error through misidentification whereas (2) is not.

We should also notice that, if we maintain that “I” is systematically ambiguous between a referring and a non-referring use, we must conclude that predicates of psychological states—such as “feel cold”—are also systematically ambiguous. Compare sentence (1), repeated below, to (3).

- (1) I feel cold
- (3) Kazimir feels cold

Sentence (1) is immune to error through misidentification. In Anscombe’s terms this means that “I” is a non-referring expression. This must also mean that the meaning of “feel cold” is such that, combined with the non-referring expression “I”, it delivers the overall meaning of sentence (1). Consider then sentence (3). It is not

immune to error through misidentification. Hence, “Kazimir” is a referring expression. This means that we must now characterize the meaning of the predicate “feels cold” in such a way that, when combined with the referring expression “Kazimir”, it delivers the meaning of (3). Ultimately, we must characterize the meaning of the predicate “feel(s) cold” in two distinct ways in the two sentences.

There is a further problem with the idea that “I” refers in (2) but does not refer in (1). Both (1) and (2) seem to support an *existential entailment*. A common intuition among speakers of English is that the truth of (2) entails the truth of (4).

- (2) I am wearing red shoes
- (4) Somebody is wearing red shoes

Notably, the same judgment applies to sentence (1), which is most naturally understood by English speakers as entailing (5).

- (1) I feel cold
- (5) Somebody feels cold

This observation raises two problems for the view that “I” is a referential expression in (2) but not in (1). First, if the meanings of (1) and (2) are so fundamentally distinct, what explains their equivalent logical properties and, in particular, the fact that they both support an existential entailment? Secondly, if “I”, as it occurs in sentence (1), does not contribute a reference, why do speakers find it so natural that (1) entails (5)? The fact that (1) entails an existential statement such as (5) suggests that natural language speakers do take (1) to entail a *de re* proposition, attributing a property to a *res*.

## Cogito Ergo Sum

There is an alternative to the view that “I” is a non-referring expression, which explains the phenomenon of immunity to error through misidentification without dismissing the idea that the first-person pronoun is a referring expression. This view maintains that “I” does refer to the individual speaking, although not as a physical object but, rather, as a disembodied mind—or, as we will call it, a *Cartesian ego*. Both Wittgenstein and Anscombe were well aware of this possibility but, as we will see, they ultimately rejected it because of its burdensome philosophical implications.

To appreciate the notion of a Cartesian ego it is useful to return to Sensory Deprived Kazimir floating in the tank. Above, we used Anscombe’s imaginary tank scenario to prove the negative conclusion that introspection does not support identification. In the tank, Sensory Deprived Kazimir is introspectively aware of what goes on in his mind. However, being deprived of any information about the outside world, including awareness of his own body, Kazimir’s introspective capacity grants him no knowledge of either the material world around him or what sort of physical object he is. In the tank, however, Sensory Deprived Kazimir is actively thinking and he is introspectively aware of doing so. Introspection may not provide him any

knowledge of who he is as a physical object but certainly allows him to identify himself as a *thinking subject*—the owner and experiencer of his own thoughts. This suggests that, even though introspection does not grant Kazimir the capacity to identify himself as a material object, it nevertheless allows him to identify himself as the thinking subject of his own mental life. After all, we may claim, introspection is a form of identification. What is peculiar about it is that the thing it identifies is of a mental nature, rather than a material one. Following the tradition set by Descartes, we shall call the entity that is identified through pure introspection a Cartesian ego.

Before discussing the broader implications of Cartesian egos, it is useful to notice that they offer a practicable solution to the problem of the implicit *de se*. By introducing cartesian egos as an independent class of objects in our model of interpretation of natural language, we can maintain that “I” refers to the speaker even when it contributes implicit *de se* judgments. What makes “I” special in such case is not that it lacks a reference, rather, that its reference is a Cartesian ego. Hence, when uttering (1) Kazimir is indeed referring to himself, although not as a material body, but a Cartesian ego—the immaterial subject of his own private experience of feeling cold.

(1) I feel cold

The fact that “I” is immune to error through misidentification, when contributing to implicit *de se* judgments, follows from the fact that Cartesian egos, being purely mental objects, can be acknowledged only in the privacy of an individual own mind and without error.

There is, of course, a heavy price to pay for this type of solution. Admitting Cartesian egos in the model of interpretation of natural language also means rejecting semantic externalism. Wittgenstein’s was already aware of this problem and, in fact, rejected the idea that “I” may refer to a Cartesian ego. In his mind, “we feel that in the case in which ‘I’ is used as subject, we don’t use it because we recognize a particular person by his bodily characteristics; and this creates the illusion that we use this word to refer to something bodiless, which, however, has its seat in our body. In fact, this seems to be the real ego, the one of which it was said, ‘Cogito, ergo sum.’” (Wittgenstein 1958, p. 69 of original division). The Cartesian motto *cogito ergo sum*—“I think therefore I am” reminds us that Cartesian egos exist only in as much as they are the product of a thinking mind. They are, in fact, the very essence of a thinking mind. As such, Cartesian egos squarely qualify as ontologically subjective entities whose use in a theory of linguistic meaning is strictly incompatible with semantic externalism.

As Anscombe made explicit in her writings, we are forced into a choice between two mutually exclusive options. Either we renounce to the view that the meaning of “I” is its reference or we give up semantic externalism. In the context of the first-person, we find that the view that meaning is reference and the view that meaning is external to the mind of speakers are just incompatible with one another. On the one hand, if we wish to maintain that the meaning of “I” is its reference, we must accept that its reference is an immaterial subject—a Cartesian ego. On the other hand, if we wish to maintain that the meaning of “I” is grounded in the external word of material



things, we must give up the possibility that its meaning is its reference. The first-person takes us to the core of subjectivity in natural language and forces us into a choice between the view that meaning is reference and the view that meaning is external to the mind.

## “I” as Expletive

Before concluding, it is worth briefly reviewing the theory of François Recanati, who recently proposed an analysis of the first-person that aims at accommodating Anscombe’s original view of the first-person as a non-referring expression within the model-theoretic account of natural language meaning that we have familiarized ourselves with in the first-part of the book. The core idea at the basis of Recanati’s analysis is that the first-person, as it occurs in implicit *de se* judgments, is an *expletive* term. That is, “I”, as it occurs in a sentence such as (1), is to be regarded as equivalent to “it”, as it occurs in a sentence such as (6).

- (1) I feel cold
- (6) It is raining

Linguists refer to the pronoun “it”, as it occurs in (6), as an *expletive subject*. By that, they mean that “it” performs a grammatical function but not a semantic one—that is, it does not contribute a meaning. On the one hand, the rules of English syntax require that all declarative sentences have an overt subject. The expletive term “it”, by combining with the predicate “is raining”, satisfies precisely this requirement. On the other hand, “it” does not contribute a meaning of its own to the sentence, as the subjects of the declarative sentences we have considered so far. Sentence (6), in effect, does not express a *de se* proposition. The sentence does not tell us that a certain *res* satisfies the property of being raining. Rather, it tells us that there is some raining going on. This proposition corresponds to a property of possible states of the world that is true of all those possible worlds that meet a certain meteorological condition—that of witnessing rain. Notice that the fact that “it” does not contribute a meaning of its own has important implications for how we understand the meaning of the predicate “is raining”. If the proposition contributed by (6) is the product of the meanings of its parts, as the principle of compositionality dictates, but “it” does not contribute a meaning of its own, then we must conclude that the predicate “is raining” does not refer to a property of objects, which would map an object into a proposition. Rather, it must contribute, in and by itself, a whole proposition—the property of all those possible worlds where it is raining.

According to Recanati, sentence (1) should be analysed in the same way as (6). Its subject—“I”—should not be regarded as contributing a *res* and, accordingly, its predicate—“feel cold”—should not be regarded as contributing a property of objects. Rather, “I” should be regarded as an expletive subject and “feel cold” should be regarded as contributing, in and by itself, a proposition.



There is, of course, an important difference between the proposition expressed by (6) and that expressed by (1). Whereas the proposition expressed by (6) corresponds to a property of *material* states of the world, which is true of all possible states of the world where it rains, the proposition contributed by (1) corresponds to a property of *psychological* states, which is true of all possible states of the mind that are characterized by the experience of feeling cold.

A common reaction to Recanati’s proposal is to wonder why a sentence such as (1) is taken to say something about the speaker. If (1) simply expresses a property that is true of all psychological states that are characterized by a feeling of cold, why do we interpret it as a judgment the speaker is making about herself and her own psychological state, not someone else’s? The answer to this question lies, again, in the parallel with (6). Sentence (6), as we saw, expresses a property that is true of all those possible states of the world where it is raining. The actual circumstances against which this property is evaluated depend on where and when the sentence is uttered. If the sentence is uttered by Kazimir in Saint Petersburg during a rainy day, the property proves true. If it is uttered by Frida during a sunny day in Cuernavaca, the property proves false. The same rationale applies to (1). Sentence (1) expresses a property that is true of all psychological states that are characterized by a feeling of cold. The actual circumstances against which this property is evaluated are determined by the occasion in which the sentence is uttered. If, for example, the sentence is uttered by Kazimir, the property is evaluated against the circumstances of his utterance—that is, Kazimir’s psychological state at the time of his utterance. If it is uttered by Frida, the property is evaluated in the context of her psychological state at the time of her utterance. Put another way, (1) is equivalent to (7).

(7) It feels cold

If we hear Kazimir uttering (7), we take him to express a property that, being a property of psychological states, is bound to concern his own subjective experience of it, not someone else’s. That is, we take (7) to express the fact that there is a feeling of cold and, since feelings are psychological and, therefore, subjective, we take it that Kazimir is inevitably talking from his own subjective perspective of experimenter of such a feeling. Ultimately, sentence (1) is about the speaker not because the speaker is the *res* the sentence is about but because the speaker provides the circumstances against which the proposition is evaluated as being true or false. Recanati’s approach gives substance, in this way, to an intuition that originates with Lewis and according to which, in first-person judgments, the first-person does not contribute a content to the judgment but, rather, provides its circumstances of evaluation.

The framework proposed by Recanati has significant advantages. To begin with, the systematic ambiguity of the first-person pronoun becomes less surprising in light of the parallel between the first-person and other known expletive terms. It is a common property of expletive terms amongst the languages of the world that they are ambiguous between an expletive use and a referential one. “It”, for example, can be used as an expletive—as in sentence (6)—as well as referentially—as in the sentence “it is on the table”, where “it” refers to an object.

With his framework, Recanati is also able to account for the existential entailment of implicit *de se* sentences. As we saw above, speakers typically infer the truth of the existential statement in (5) from the truth of (1).

- (1) I feel cold
- (5) Somebody feels cold

But how are speakers able to conclude that there is someone who feels cold from (1), if the meaning of (1) does not entail reference to any object at all? According to Recanati, the inference from (1) to (5) is supported by a process he calls “reflection”. Reflection is a process of transition from a premise to a conclusion that requires no evidence except the truth of the premise and what Recanati refers to as the “mode of the grounding experience”—the experience, that is, that grounds the subject’s judgment that the premise is true. Again, the parallel with expletive terms helps us clarify this notion. Consider the inference from (6) to (8).

- (6) It’s raining
- (8) It is raining here

Suppose that Kazimir utters (6) after he has come to realize through a sensorial experience—say, by looking out of his window—that it is raining. Given its indexical nature, perception is bound to concern the place and time where the perceiver is. Hence, Kazimir can safely conclude from (6) that (8) also holds true. This inference is based on nothing else but the truth of (6) and the properties of the sensorial experience that supported Kazimir’s judgment in the first place. In a sense, (8) can be regarded as expressing a sort of meta-proposition—that is, a proposition about a proposition. Whereas (6) expresses the impersonal proposition that there is an event of raining taking place, (8) expresses the *de re* proposition that the proposition expressed by (6) is true of the current state of the world. The same process of reflection can be taken to support the inference from (1) to (5). When Kazimir acknowledges the truth of (1), he does so through his own introspective experience of cold. This experience also grants him knowledge that he is a subject who is experiencing such experience. Hence, through reflection, Kazimir can infer the truth of (5) from (1). As (8), also (5) can be regarded as the outcome of realizing the meta-proposition that the proposition expressed by (1) is true of the person speaking.

Recanati’s approach delivers important results but also meets some objections. From a linguistic perspective, it can be observed that the parallel between the first-person and expletive terms holds only to a certain extent. Across different languages, we find that not all first-person sentences have a corresponding expletive variant, in the way (1) corresponds to (7), repeated below.

- (1) I feel cold
- (7) It feels cold

In English, for example, we find that most predicates of emotion or perceptual experience do not have an equivalent parallel. Sentences (9) and (10), to see a concrete

case, do not have equivalent impersonal variants such as “it feels happy” or “it hears trumpets”.

(9) I feel happy

(10) I hear trumpets

In the context of our current discussion, there is also a more conceptual concern with Recanati’s framework. The framework Recanati proposes, although very valuable in practice, does not really help us in addressing Anscombe’s critical conundrum. It does not help us, that is, in reconciling the meaning of the first-person with the principles of semantic externalism. Indeed, within his framework, Recanati is able to implement Anscombe’s idea that the first-person does not refer to an object, at least when contributing to implicit *de se* judgments. However, the framework does not really allow us to avoid reference to objects that are subjective in the ontological sense. Central to Recanati’s framework, in fact, is the distinction between two classes of properties: properties of material states of affairs, on the one hand, and properties of psychological states, on the other. This distinction, as we saw above, is indispensable if we wish to account for the first-personal, self-oriented nature of the implicit *de se*—the fact that, when we hear Kazimir saying “It feels cold”, we automatically understand it as expressing a psychological state of which Kazimir himself is the experiencer. It is precisely because the predicate “to feel cold” expresses a property of psychological states, which can be contemplated only in the privacy of the speaker’s own mind, that it is also automatically understood as concerning the speaker and not somebody else. Introducing the distinction between material and psychological properties, however, also means introducing a corresponding distinction between two different classes of entities in the model of interpretation—objective entities and subjective ones. Properties, as we know, are equivalent to the set of objects of which they are true. A proposition—conceived as a property of material state of affairs—is equivalent to the set of possible worlds of which it is true. The proposition expressed by the sentence “Kazimir is a painter”, for example, corresponds to the set of all possible worlds where Kazimir is a painter. By the same token, a property of psychological states is but the set of all possible states of an individual’s mind of which the property is true. Hence, the property expressed by the sentence “I feel cold” is equivalent to the set of all psychological states that are characterized by a feeling of cold. This means that psychological states are now one of the essential classes of objects in our model of interpretation. Ultimately, Recanati dispenses us from the need to directly refer to Cartesian egos but does so by replacing reference to Cartesian egos with reference to their properties. Properties of Cartesian egos, however, are as incompatible with semantic externalism as Cartesian egos themselves.

On a related note, we should also notice that Recanati’s process of reflection—as applied to implicit *de se* judgments—is, after all, but another incarnation of the Cartesian *cogito ergo sum*. Reflection, as we saw a moment ago, is the process that allows us to conclude that “somebody feels cold” from the fact that “it feels cold”. The process of reflection, Recanati contends, is based solely on the truth of the premise—the fact that it feels cold—and the mode of the experience that grounds

the judgment that the premise is true. In the case of implicit *de se* judgments, however, the mode of the grounding experience is introspection. If reflection is to achieve its goal, then, it must consist in a process that allows us to conclude that we exist, as the subjects of our own experience of cold, on the basis of our introspective experience of cold alone. This is but an incarnation of the Cartesian scheme “I think, therefore I am”.

Recanati’s framework adopts Anscombe’s practical insight that the first-person is a non-referring term but does not exploit it to address Anscombe’s philosophical concerns. Ultimately, the dilemma raised by Anscombe stands strong. Any attempt at explaining the meaning of the first-person on the basis of the notion of reference is bound to contradict the principles of semantic externalism. Either we abandon the view of meaning as reference—provided we are ready to face the consequences of such choice—or we abandon the principles of semantic externalism. In the following chapter, we will explore the latter option.

## References and Remarks

Skeptical arguments of the Cartesian sort have been formulated in different forms and across different cultures. A notable example is the classic Chinese poem “Zhuang Zhou Dreams of Being a Butterfly”, here reported in the Watson translation (Watson 1964):

Once Zhuang Zhou dreamed he was a butterfly, a butterfly flitting and fluttering around, happy with himself and doing as he pleased. He didn’t know he was Zhuang Zhou.

Suddenly he woke up, and there he was, solid and unmistakable Zhuang Zhou. But he didn’t know if he were Zhuang Zhou who had dreamed he was a butterfly or a butterfly dreaming he was Zhuang Zhou. Between Zhuang Zhou and a butterfly, there must be *some* distinction! This is called the Transformation of Things.

The poem is regarded as one of foundational texts of Daoism. It describes a man who dreams of being a butterfly and, upon waking up as a man, doubts whether he is a man or a butterfly. A more recent version of the Cartesian skeptical argument is the so-called brain in a vat argument, originally due to Gilbert Harman (Harman 1973). We are asked to contemplate the possibility that, instead of being human beings endowed with a body and living in the material world we experience in our daily lives, we are disembodied brains kept in a vat of nutrients by a super computer. Our experiences of our bodies and the material world around us are induced by the supercomputer by sending especially designed electro-chemical charges to our disembodied brains. As in the case of the Cartesian argument, the challenge resides in demonstrating that we are not brains in a vat (some readers may have recognized in this scenario the philosophical premises of the movie *The Matrix*). A rebuttal of the brain in a vat argument was attempted by Putnam (Putnam 1981). Putnam’s counterargument is relevant in the context of our discussion because it is based on the notion of semantic externalism. We will illustrate the argument very briefly and in a

simplified form. Semantic externalism requires that the meaning of the words we utter be causally determined by the external reality in which we utter them. If the thesis applies to all words, then it must also apply to the words “brain” and “vat”. If we were brains in a vat, there will be no external reality causally determining the meaning of words such as “brain” and “vat”. Hence, the very fact that we can utter and understand a sentence such as “I am a brain a vat” is evidence of the fact that we are not brains in a vat. There is a rich, largely critical, literature on Putnam’s argument against the brain in a vat scenario. From our perspective, it is worth pointing out the following potential flaw in Putnam’s strategy. Putnam’s argument is that, if semantic externalism is correct, then the brain in a vat scenario is false. However, one could very well argue that it is the very impossibility of rejecting the brain in a vat scenario that provides evidence that semantic externalism is incorrect.

The notion that the first-person is a non-referring term is grounded in a philosophical tradition whose most notable champion is David Hume. In his *Treatise of Human Nature* (Book I, IV, vi), Hume famously claims:

There are some philosophers who imagine we are every moment intimately conscious of what we call our self; that we feel its existence and its continuance in existence; and are certain, beyond the evidence of a demonstration, both of its perfect identity and simplicity. The strongest sensation, the most violent passion, say they, instead of distracting us from this view, only fix it the more intensely, and make us consider their influence on self either by their pain or pleasure. To attempt a farther proof of this were to weaken its evidence; since no proof can be derived from any fact, of which we are so intimately conscious; nor is there any thing, of which we can be certain, if we doubt of this.

Unluckily all these positive assertions are contrary to that very experience, which is pleaded for them, nor have we any idea of self, after the manner it is here explained. For from what impression could this idea be derived? This question it is impossible to answer without a manifest contradiction and absurdity; and yet it is a question, which must necessarily be answered, if we would have the idea of self pass for clear and intelligible. It must be some one impression, that gives rise to every real idea. But self or person is not any one impression, but that to which our several impressions and ideas are supposed to have a reference. If any impression gives rise to the idea of self, that impression must continue invariably the same, through the whole course of our lives; since self is supposed to exist after that manner. But there is no impression constant and invariable. Pain and pleasure, grief and joy, passions and sensations succeed each other, and never all exist at the same time. It cannot, therefore, be from any of these impressions, or from any other, that the idea of self is derived; and consequently there is no such idea.

Since Hume, the idea there is no such thing as the self—the immaterial subject of conscious experience—has found many different incarnations and linguistic applications such as those of Anscombe (1975) and Recanati (2007, 2012), which we reviewed in this chapter. Notably, the view that immunity to error through misidentification is not the product of identification through introspection but, rather, of the fact that introspection does not involve identification at all is also the view promoted by Shoemaker. This view is most clearly expressed in lecture I of his “Self-Knowledge and ‘Inner Sense’” (Shoemaker 1994, pp. 257–258):

In introspective self-knowledge there is no room for an identification of oneself, and no need for information on which to base such an identification [...] There are indeed cases of

genuine perceptual knowledge in which awareness of oneself provides identification information, as when noting the features of the man I see in the mirror or on the television monitor tells me that he is myself. But there is no such role for awareness of oneself as an object to play in explaining my introspective knowledge that I am hungry, angry, or alarmed. This comes out in the fact that there is no possibility here of a misidentification; if I have my usual access to my hunger, there is no room for the thought “Someone is hungry all right, but is it me?”

The literature also offers attempts at defending intermediate positions that maintain introspection is a form of identification but try to do so without having to endorse a Cartesian metaphysics of the self. Notable examples are Evans (1982) and James (1976).

It should also be observed that, quite independently of its metaphysical commitments, the notion of a self that is the immediate subject of conscious experience and is distinct from the self as the physical object has played a central explanatory role in a number of frameworks in cognitive science, at least since William James’s distinction between the notions of *I*—the self as subject—and *ME*—the self as object (James 1890). A recent example is Shaun Gallagher’s distinction between “minimal self” and “narrative self” (Gallagher 2000).

Finally, we should observe that Higginbotham’s theory of the implicit *de se* in control structures (Higginbotham 2003), which we briefly reviewed in the references and remarks section of the previous chapter, also assumes an element of ontological subjectivity. Core to Higginbotham’s proposal, in fact, is the idea that the subject of the embedded clause is identified as the subjective experiencer of the event expressed by the main predicate.

## References

- Anscombe, E. (1975). The first person. In S. D. Guttenplan (Ed.), *Mind and language* (pp. 45–65). Oxford: Oxford University Press.
- Evans, G. (1982). *The varieties of reference*. Oxford: Oxford University Press.
- Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, 4, 1.
- Harman, G. (1973). *Thought*. Princeton, NJ: Princeton University Press.
- Higginbotham, J. (2003). Remembering, imagining, and the first person. In A. Barber (Ed.), *Epistemology of language* (pp. 496–533). Oxford: Oxford University Press.
- James, W. (1890). *The principles of psychology*. New York: Holt.
- James, W. (1976). Does ‘consciousness’ exist? In W. James (Ed.), *Essays in radical empiricism* (pp. 3–19). Cambridge, MA: Harvard University Press.
- Putnam, H. (1981). Brains in a vat. In *Reason, truth, and history* (pp. 1–21). Cambridge: Cambridge University Press.
- Recanati, F. (2007). *Perspectival thought: A plea for (moderate) relativism*. Oxford: Oxford University Press.

- Recanati, F. (2012). Immunity to error through misidentification: What it is and where it comes from. In S. Prosser & F. Recanati (Eds.), *Immunity to error through misidentification: New essays* (pp. 180–201). Cambridge: Cambridge University Press.
- Shoemaker, S. (1994). Self-knowledge and ‘inner sense’. *Philosophy and Phenomenological Research*, 54, 2.
- Watson, B. (1964). *Zhuangzi: Basic writings*. New York, NY: Columbia University Press.
- Wittgenstein, L. (1958). *Preliminary Studies for the “Philosophical Investigations”, Generally known as The Blue and Brown Books*. London: Basic Blackwell.

## Chapter 30

# Meaning and Mind



It is time to take stock. We began part II of this book by reviewing the thesis of semantic externalism—the thesis, that is, that natural language meaning is grounded in the external world of natural things and not in the mind of the speakers who entertain it. In the course of part II, we also considered three main objections to semantic externalism.

The first is that the logic of Wittgenstein’s argument against private language applies also to public language, with equally disruptive consequences. After closer inspection, we find that the rules of public language are never fully deterministic—that is, they are never truly capable of predicting all their potential future uses. Whether a newly encountered object deserves to be called a “table” depends on a stipulation—an arbitrary decision on the part of speakers—settling whether the object does or does not qualify as a table. If we were to follow the logic of Wittgenstein’s argument to the letter, we should then conclude that also public language is a moot exercise—a practice that cannot be described as the unfolding of a pre-established set of rules. As we saw, a distinctive reading of Wittgenstein maintains that this is not really an objection to Wittgenstein’s private language argument but, in fact, the more general lesson we should draw from it. Language—private as well as public—should not be described as the unfolding of a pre-established set of a priori rules. It should rather be understood as a behaviour—a set of actions whose value depends solely and exclusively on the effect they produce a posteriori on their environment. We observed, however, that this view of language is incompatible with the way speakers learn language. Contrary to what was generally believed at the time when Wittgenstein’s was formulating his argument, language is not learned on the basis of sole interaction with the environment, as it should be expected in light of the view of language as behaviour. Learners do not learn language by trial-and-error—that is, by attempting linguistic behaviours and evaluating a posteriori their success on the basis of the feedback they receive from their environment. Rather, learners approach language by relying, from the very start, on the support of a rich system of inherited cognitive principles. These facts support the view that the way



speakers learn and, then, use language is, at least in some measure, the product of a set of pre-determined rules and principles.

The second objection we considered against semantic externalism is that the way speakers talk about things is not always equivalent to the way things are. Semantic externalism predicts that the way speakers talk about things follows straightforwardly from the way things are in the external world of natural things. Against this prediction, we observed that meaning often comprises an abstract dimension. When talking about “books”, for example, speakers do not only mean corresponding material objects but also abstract contents. Similarly, when talking about “cities”, speakers do not only mean soil, buildings, and people but also history, traditions, and institutions. In this respect, we also observed that the type of abstraction we find in linguistic meaning is quite different from the one we find in the interpretation of formal languages such as mathematics and, therefore, cannot be explained away in the same manner. As we saw, a number of philosophers, beginning with Plato, have defended the view that mathematical entities have an objective status, despite being immaterial. These philosophers maintain that the world comprises both a concrete and an abstract dimension— a realm of Platonic ideas that are immaterial but, nonetheless, real—and that the purpose of mathematics is to describe such abstract dimension. This explanation, we observed, does not work for the natural language. If we took the abstraction of books and cities to belong to the realm of Platonic ideas, we would find ourselves facing an intractable epistemic problem: How do speakers manage to acknowledge, entertain, and learn about such objects?

The third and final objection we considered against semantic externalism is that there is an irreducible core of ontological subjectivity in natural language meaning. The evidence for this objection comes from implicit *de se* judgments, which are those judgments speakers make about themselves in the first-person and without the possibility of error. As we saw in the last chapter, the problem of self-reference in introspective judgments takes us to a fundamental opposition between two radically different and mutually exclusive options: either abandoning the view that meaning is reference or introducing in the model of interpretation of natural language entities that are ontologically subjective—such as Cartesian egos. The first option—Wittgensteinian in spirit—encounters both conceptual and practical objections. The second option contradicts semantic externalism. The lesson we drew is that, if meaning is what language is about, then language is about, among other things, ontologically subjective entities.

As the attentive reader may have already noticed, the three objections are connected by a common conceptual thread. In all three cases, semantic externalism appears to be missing the essential qualities of the cognitive relation between meaning and its users—the relation that enables them to learn meaningful language and to use it to express abstractions as well as inner thoughts, feelings, and emotions. The core principle of semantic externalism is that meaning is fully understood in the relation between language and the external world. By this very definition, semantic externalism simply lacks the conceptual structure to explain how linguistic meaning plays such an integral part in the cognitive lives of its users.

## Weak Semantic Externalism

To be sure, very few scholars would agree today with the thesis of semantic externalism in the strong form we have maintained so far. Up to this point, we have understood semantic externalism as the thesis the *all* meaning of *all* natural language expressions is grounded in the external world of natural things. This stronger version of semantic externalism has had the support of eminent thinkers since its earliest formulations. Yet, many today would agree with a weaker interpretation of the thesis, whereby all that semantic externalism requires is that *some* of the meaning of *some* natural language expressions is grounded in the external world of natural things. This weaker version of the thesis allows for the possibility that the meaning of some expressions is, at least in some part, not grounded in the external world of natural things. Weak semantic externalism tells us that a satisfactory explanation of linguistic meaning must not neglect its link to the external world while leaving room for other ontological primitives to take part in its definition.

What is unsatisfactory about this weaker version of semantic externalism is that it offers an incomplete answer to the question of what is meaning. The thesis tells us that a part of linguistic meaning is grounded in the world outside the mind of speakers but fails to tell us what we should make of the other part—the one that is not grounded in the external world. What are we to make of this aspect of linguistic meaning? Are we legitimated to conclude that meaning is also a psychological object? But, then, what about Wittgenstein's and Putnam's arguments against this possibility?

## Semantic Internalism

Each of the three main arguments against semantic externalism we have reviewed in the course of part II is also an argument for the opposite thesis: *semantic internalism*. This is the thesis that meaning is a mental object. According to it, meaning is not a function of the world but of the mind of the speakers. Language is a means to express mental contents—thoughts, concepts, ideas—rather than reference to external objects. It is a thesis of that has had eminent defenders, from Aristotle to Locke.

Of course, the most significant advantage of semantic internalism is that it naturally captures the cognitive role meaning plays in the psychological lives of speakers. How speakers manage to acknowledge, master, and entertain meaning is no longer an impossible epistemic challenge but a natural consequence of the fact that meaning is a psychological entity. Also, in the framework of semantic internalism, it is no longer surprising to find that meaning comprises an abstract dimension. If correct, semantic internalism entails that speakers do not talk about books and cities as external objects whose properties hold irrespectively of the way they think about them but, rather, as mental objects—the products of a mental activity. By the same

token, semantic internalism naturally explains how speakers are able to use language to express their inner thoughts, feelings, and emotions.

Aside from these advantages, however, semantic internalism also suffers from fundamental problems. A first problem is that, if meaning is a psychological object, rooted in the inner realm of the mind, then it is also a private one. Obviously, this conclusion stands in sharp contrast with Wittgenstein's private language argument, which contends that meaning, whatever it may be, cannot be private. We should observe that, in effect, a private understanding of meaning, such as the one semantic internalism promotes, renders the whole notion of reference a vacuous one. As we already saw in Chap. 6, the core idea at the foundation of the thesis that meaning is reference is that language provides a *homomorphism* between two domains of objects: the expressions of the language, on the one hand, and the objects these expressions are about, on the other. With his argument, Wittgenstein warns us that, to avoid turning this logical structure into a vacuous exercise, we must ensure that the two sides of the homomorphism come from distinct and independent sources. In spite of this warning, semantic internalism does just that. According to it, both the language and its domain of reference are rooted within one and the same realm—the psychological life of the speaker. In such a framework, it hardly makes sense to speak of meaning in terms of reference. As the structuralist theory and the theory of meaning as use, semantic internalism ends up obliterating the distinction between language and its meaning and, ultimately, qualifies as a theory of meaning as nothing.

Regarding meaning as a private object is problematic for a further reason. Meaning is something speakers communicate to one another. True, there are many circumstances in which language is used for private purposes, for example, when taking notes on a personal diary or when thinking in language. This, however, does not affect the obvious observation that meaning is something speakers are capable of communicating to one another. How can meaning be communicated from a speaker to another unless it is something that can be shared publicly? Grounding meaning in the external world is important not only to explain its causal history but, indeed, also to provide a public domain where it can be shared by different speakers.

Another challenge to semantic internalism is provided by Putnam's Twin-Earth experiment. To best appreciate the implications of Putnam's experiment for the thesis of semantic internalism, it is useful to briefly rehearse its basic structure. In his thought-experiment, Putnam compares two identical speakers—Kazimir and Twin Kazimir—in two identical environments—actual Earth and Twin Earth. The two environments differ only and exclusively in one feature: The liquid, colorless, and transparent natural substance that runs in rivers, fills lakes and oceans, and is sold in bottles is H<sub>2</sub>O on actual Earth but a different chemical compound on Twin Earth. Water on actual Earth and water on Twin Earth, that is, have the same superficial features—they are both a liquid, colorless, and transparent natural substance that runs in rivers, fills lakes and oceans, and is sold in bottles—but are, ultimately, two different natural objects. Putnam observes that, when actual Kazimir and Twin Kazimir utter the word "water", they end up talking about different things and this is so irrespectively of the fact that they have the very same thoughts about the natural

substance they drink and wash themselves with. Remember that, in the thought experiment, actual Kazimir and Twin Kazimir are identical down to the smallest molecule, which means that they also share the very same brain and, with it, the very same thoughts. This demonstrates that the meaning of the word “water” depends exclusively on what water is and not on what the two Kazimir’s think it is.

Putnam’s conclusion is obviously impossible to reconcile with the thesis of semantic internalism. If semantic internalism were correct, we would expect that the word “water” receives the same meaning when in the mouth of the two Kazimir’s. If, that is, the meaning of the word “water” depended solely and exclusively on what the speakers think of it, we should conclude that actual Kazimir and Twin Kazimir express the very same meaning when uttering the word “water” because, in Putnam’s scenario, actual Kazimir and twin Kazimir share the same brain and, therefore, the exact same thoughts about what water is. This, however, is not the intuition we have when confronted with Putnam’s scenario. Our intuition is, rather, that the two Kazimir’s talk about different things when uttering the word “water” and, therefore, express different meanings.

Of course, there is a sense in which, in effect, actual Kazimir and Twin Kazimir do mean the same thing when uttering the word “water”. If, say, actual Kazimir were to travel to Twin Earth and ask for a glass of “water”, what he would get in exchange would be a glass of the liquid, colorless, and transparent natural substance that runs in rivers, fills lakes and oceans, and is sold in bottles. He would drink it and find it as refreshing as the water on actual Earth. In fact, he could very well live the rest of his life on Twin Earth without ever noticing that what he is talking about when saying “water” is not the same thing as what he called “water” on actual Earth. The same would happen, indeed, if Twin Kazimir were to visit actual Earth. He would obtain what he wants when asking for “water” and, similarly, actual Earth speakers of English would understand what he means when uttering the word “water”. This suggests that, whereas it may be true that “water” refers to different natural objects in the two planets, it does nonetheless mean the same thing to the two Kazimir’s—at least when it comes to the practical sphere of their goals and intentions.

This may very well be true, but it is not enough to prove that the word “water” has the same meaning on the two planets. This is demonstrated by the simple fact that one and the same sentence is true on one planet—actual earth—but false on the other—twin earth:

(1) Water is H<sub>2</sub>O

Whereas the two Kazimir’s may think of the meaning of the word “water” in exactly the same way and, to some extent, be able to perform the same functions by using it, the fact that (1) is true in one planet but false in the other shows us that “water” contributes different propositions in the different planets and, therefore, expresses different meanings. Putnam’s experiment is, after all, a formidable argument against semantic internalism.

## Weak Semantic Internalism

As in the case of semantic externalism, very few scholars would agree today with the strong version of semantic internalism we have just discussed. Most advocates of semantic internalism would rather side with a weaker version of the thesis, which does not deny that speakers sometimes use language to refer to the external world but also maintains that this is not really relevant when it comes to providing a theory of linguistic meaning. According to this version of semantic internalism, the goal of a theory of linguistic meaning is that of providing an account of the cognitive skills that allow speakers to speak meaningfully. This goal, it is claimed, does not require taking into account the actual use that speakers make of these skills, which may include, among other things, reference to external objects. This is the view most notably endorsed by Noam Chomsky.

To better appreciate this form of semantic internalism, it is useful to consider an analogy with another human skill—the capacity for movement. Differently from plants, and similarly to other animals, humans have the capacity to move in their environment. Let us imagine for a moment that we were scientists interested in explaining this capacity. What should our explanation look like to be regarded as a satisfactory account of the human capacity for movement? Most likely, it would entail a description of the human motor system, including its mechanics—how bones and muscles work together to enable movement—as well as its cognition—how the brain plans, initiates, and controls movement. Crucially, however, we would not be expected to include in our explanation what humans actually do with this capacity. Nor should we bother about the social norms or conventions that regulate its use. Our account, in other words, should explain how humans move regardless of whether they use to have a nice stroll in the park or to go to work. As irrelevant to our explanation would be the fact that social norms dictate that one should not walk across the street with a red light. All that really matters is that we offer a complete description of the internal states that allow humans to perform all of these actions.

The same—advocates of weak semantic internalism submit—should go for the capacity to speak meaningful language. Our explanation of the human capacity to speak meaningful language should include a description of the physiological and cognitive means that allow competent speakers to express meaning through language. Crucially, however, what speakers do with this capacity, to which ends they employ it, and what norms regulate its use should all be regarded as factors irrelevant to our explanation. Ultimately, speakers may very well use language in reference to the external world and its objects. Furthermore, their use of language may be regulated by all sorts of public norms and conventions. Yet, these facts are all irrelevant when it comes to describing the internal capacity of speakers to learn and use language meaningfully, which is what a theory of linguistic meaning should really be about.

This weak version of semantic internalism is less stringent than the one we have considered above. It does not exclude that language may be used to refer to the external world and does not exclude that language may have a social function and

be shaped by social norms and conventions. Yet, it also maintains that all these aspects are irrelevant to a theory of language, which should be focused on describing those internal states—physiological and cognitive—that allow speakers to speak.

As it was the case with weak semantic externalism, also this weaker version of semantic internalism is ultimately unsatisfactory. Perhaps, its most significant weakness is that, by focusing exclusively on the internal states of language users, it fails to tell us what factors shaped the human capacity to use language meaningfully in the form we observe it today. Once again, this point can be best appreciated by comparing the capacity for language to the capacity for movement. Let us suppose that we now have a comprehensive description of the human motor system—its physiology as well as its cognition. Suppose we now want to investigate also the reasons why the system has the properties it has. We want to understand why, of all possible systems of locomotion, humans have ended up with the one they have today. Darwin's theory of evolution suggests that the factors that shaped human motor skills lie in the evolutionary history of the human species. At some point in the evolutionary history of the human beings, these skills must have been produced by a mutation, they must have survived, and they must have been transmitted to the next generations. But why these skills and not others? What made these skills so special that they were not discarded in the course of human evolution, as it is the fate for the vast majority of genetic mutations? If they survived, Darwin tells us, it is because they served humans well or, at least, they did not harm them. This answer, of course, raises a further question: What determined the usefulness of a skill and, with it, its survival? Darwin's answer is that how useful a skill is depends on the environment in which it is performed and the use it is put to. The human motor system evolved in the way it did because it allowed humans to perform useful tasks in the environmental niche in which they found themselves. If humans had evolved in a different environmental niche, their motor system would have developed differently. If, for example, humans had evolved in an environment characterized by a different gravitational force, our legs and feet would look very different today or, perhaps, we would not even have legs and feet but an altogether different system of locomotion. In the end, an explanation of the factors that determine the shape of an organism's skill—such as the ability to move—necessarily include the environment where it is performed.

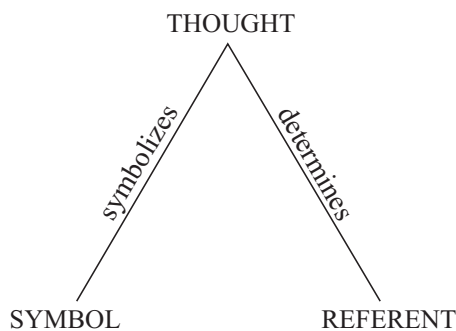
The same rationale can be applied to the capacity for meaningful language. If such capacity has evolved in the way it did, it must have been because, at some point in the evolutionary history of the human beings, it has served a useful purpose in the environmental niche in which it was put to use. Hence, if we want to properly explain the human capacity for meaningful language, we cannot disregard the function it performs in the environment in which it is realized. The weaker version of semantic internalism we have discussed in this section does just that. It allows us to describe the capacity for language as we observe it today but, by excluding from the explanation both the function meaningful language performs for its users and the environment in which it performs it, does not allow us to identify the environmental reasons and factors that shaped its current properties.

## Sense and Reference

The lesson to be drawn from our discussion of semantic externalism and semantic internalism seems clear enough: A satisfactory account of linguistic meaning must comprise both its relation to the human mind as well as that to the external world in which it occurs. Focusing on one aspect while disregarding the other is bound to deliver an incomplete account of meaning and its properties. Ultimately, the receipt for meaning must comprise both aspects amongst its essential ingredients.

The view that meaning comprises both an internal dimension—oriented towards the mind—and an external dimension—oriented towards the external world—has been contemplated in different occasions in the history of linguistic thought and by different scholars. One of the most well-known examples is provided by Ivor Armstrong Richards and Charles Kay Ogden who, in their 1923 book “The Meaning of Meaning”, propose that meaning is the product of the relation between two distinct elements: a *thought* and a *referent*. The peculiar relation between thought and referent is commonly exemplified by a triangle, as in Fig. 30.1. At the bottom-left corner of the triangle we find the linguistic expression. The linguistic expression is connected to the thought—on the top corner—by a relation of *symbolization*. The expression, that is, *symbolizes* the thought. In turn, the thought *determines* the referent of the expression. Notably, there is no direct link between the expression and its referent. Their relation is always and necessarily mediated by the thought.

A similar idea has often been attributed to Frege. In his 1892 article “Über Sinn und Bedeutung”, Frege famously distinguishes between two types of meaning, to which he refers in the original German text as *Sinn*—commonly translated as *sense*—and *Bedeutung*—typically translated as *reference*. In effect, Frege’s distinction resembles that of Ogden and Richard in a number of fundamental ways. According to Frege, the sense of a linguistic expression is the thought that is



**Fig. 30.1** A triangle representing the relation between a linguistic symbol a thought and a referent, as it was conceived by Armstrong Richards and Charles Kay Ogden in their 1923 book “The Meaning of Meaning”. The relation between symbol and thought is one of symbolization (the symbol symbolizes a thought). The relation between thought and referent is one of determination (the thought determines a referent). Notice that there is no direct relation between symbol and referent. Reference is always mediated by thought



associated with it, whereas its reference is the actual object the expression is about. Furthermore, sense *determines* reference. The sense of an expression, that is, is the abstract thought, concept, or idea that allows speakers to identify what the expression is about in the actual circumstances in which the expression is used.

To be sure, Frege's distinction between sense and reference was not originally meant as a framework to reconcile world and mind in linguistic meaning. It was meant to solve other problems. One of them is Frege's puzzle, which we already discussed in Chap. 23. The second is the problem of fictional names, such as "Sherlock Holmes" and "John Watson", which lack a reference though they do not lack a meaning. Indeed, in the context of our current discussion, it is tempting to put Frege's distinction to serve a more ambitious task—that of capturing the relation between world and mind in linguistic meaning, by regarding sense as the aspect of meaning that is relevant to the mind and reference as the one that is relevant to the external world.

Doing so, however, is problematic for a number of reasons. To begin with, it is questionable whether Frege himself would have agreed with such enterprise. Frege was a committed Platonist who regarded senses as abstract objects—akin to platonic ideas—and would have firmly objected to the view that they are psychological entities. A further, more substantial problem is that treating senses as psychological entities undermines the whole idea that sense determines reference. As we saw, Frege maintains that the reference of an expression is a function of its sense. Sense is the abstract concept or idea associated with an expression that determines the object the expression is about in the real world. If senses are psychological, however, this framework cannot be maintained. Putnam's thought experiment shows us that the psychological state speakers associate with a linguistic expression is not what determines its reference. Actual Kazimir and Twin Kazimir have identical brains and, therefore, identical psychological states when uttering the word "water". Yet, they still end up referring to different objects. Whatever it is that the two Kazimir's think of water, it is not what determines the reference of the word "water". If the sense of "water" is what determines its reference, then it cannot be a psychological state because psychological states, Putnam's thought experiment tells us, fail to do just that. In fact, Putnam's thought experiment shows us that the reference of an expression is determined by its environment of use, not by a psychological state. Senses, henceforth, are better described as *characters*, which, as we saw in Chap. 23, are functions mapping the context of utterance of an expression to its content—the object it refers to. This is, in effect, the strategy we have adopted back in Chap. 23 to solve Frege's puzzle.

Some—especially the philosopher Jerry Fodor—have reacted to this conclusion by pointing out that there is nothing wrong in regarding characters as descriptions of psychological states. After all, all Putnam's argument does is to demonstrate that senses are *relational*—they do not determine the content of linguistic expressions in and by themselves but, rather, as functions of their circumstances of employment. The argument, hence, tells us something about the logical structure of senses—the fact that they are relational—but nothing about their ontological status—whether they are natural objects or psychological ones. So, the fact that senses are akin to



characters is not at all incompatible with the view that senses are, after all, descriptions of mental states. Under this view, senses are like character in that they relate a context to a content but they are also psychological entities in that the relation of mapping a context onto a content is what minds perform.

To this view, others—most notably Robert Stalnaker and Ned Block—have objected that characters are poorly suited to represent the aspect of linguistic meaning that is relevant to the cognitive life of speakers. In effect, it is not so easy to accept the idea that what speakers grasp about the meaning of “water” is a function that, when applied to a context, delivers a reference. It is precisely because of their relational nature that it is difficult to regard characters as meaningful objects. Characters are functions that *provide* a meaning. In and by themselves, however, they are not really meaningful. A further challenge to Fodor’s view that psychological states are equivalent to characters is provided by the implicit de se. As we saw in the previous chapters, if “I” refers to anything in the implicit de se, it refers to a Cartesian ego. If this is correct, the implicit de se is problematic for Fodor’s view in at least two respects. Firstly, in the implicit de se, “I” seems to have no character at all. Its content is not a function of its environment of use. To the contrary, the content of “I” is individuated internally to the mind of speakers, through introspection. Secondly, what is psychological about the meaning of “I” in the implicit de se is not its character—if there is one at all—but its content. What is ontologically subjective about the meaning of “I”, in other words, is not the function that determines its referent, but the referent itself—the peculiar type of object “I” refers to, if, of course, it refers to anything at all.

The debate on how the worldly and cognitive dimensions of linguistic meaning can be reconciled with one another is extremely complex and, as of today, remains the object of heated debates. The challenge we face is clear enough. If we want to understand what linguistic meaning is, we must find a way to reunite, within a common general framework, the side of it that is oriented toward the external world with the side of it that is oriented towards the inner realm of the mind. The facts and arguments we have reviewed throughout this part of the book suggest that the essential difficulty in addressing our challenge stems from a shared original sin—that of endorsing, right from the start, the Cartesian view that the divide between world and mind is ultimately irreparable. Once this distinction is endorsed, no amount of work seems enough to provide a principled explanation of how these two notions manage to relate to one another in linguistic meaning. We, the authors, believe that the mistake consists precisely in endorsing the Cartesian divide. To address our challenge, hence, we must fundamentally reassess the very distinction between world and mind—a task we will undertake in the following and last part of this book.

## References and Remarks

Chomsky’s internalist views have characterized his approach to language since his first criticism of Skinner’s behaviorist theories but are most explicitly formulated in his essay *New Horizons in the Study of Language and Mind* (Chomsky 2000).

There are at least two recent approaches to natural language meaning that explicitly maintain an internalist stance. The first is the framework developed by Gillian Ramchand in her *Situations and Syntactic Structures* (Ramchand 2016). In Ramchand's framework, language expressions are assumed to refer to mental conceptual representations. In particular, reference to events plays a central role. The use of linguistic predicates is based on the capacity of abstracting and generalizing over actual instantiations of physical events. This capacity is based on the recognition of causal and dynamical relations between the actants of concrete events, in a way that abstracts away from time and place. As a consequence, the temporal structure of predicates is encoded outside the syntactic space in which the essential relational properties among the actants of the event are represented. The event concepts corresponding to the language predicates are thus the perceptual/cognitive acts by means of which properties are ascribed to individuals and relations among the actants in an event are recognized (for an analysis of propositions along similar lines in the philosophical literature, see especially Soames 2015, to which we will return at the end of part III). According to this view, the lexical items used by a speaker refer to abstract event properties which are independent of time and space. In this sense, lexical items are themselves objects in the world, bundles of form and meaning whereby meaning involves the activation of an atemporal layer of essential properties that are entirely independent of temporal or world parameters. The second approach advocating internalism is the framework developed by Paul Pietroski in his *Conjoining Meanings* (Pietroski 2018; see also Pietroski 2005). According to Pietroski, lexical items, and in particular language predicates, do not refer to extensions. Rather, each atomic meaning encodes instructions for accessing a whole set of monadic and dyadic concepts. In this sense, the polysemy of linguistic predicates becomes their most essential feature. For instance, there is no sense according to which the meaning of "water" is provided by stuff external to the mind that has a certain chemical composition, since coffee contains more H<sub>2</sub>O than certain samples of the stuff that are correctly referred to as "water" (as traditionally held by Chomsky). Similarly, the abstract and physical meanings of "book" are not defined by different extensions; rather, the word "book" is a complex instruction to activate a complex network of inherently related concepts. As for non-atomic meanings, that is, the meanings that are associated with complex linguistic expressions like phrases and sentences, Pietroski argues in a parallel anti-extensionalist vein that there is no hope for the classical position that this non-atomic meaning should be identified with the truth-conditions allegedly associated with sentences. Analogously to Ramchand (and, as we will see in part III, Soames), Pietroski does not deny that meanings can be associated to truth-evaluable thoughts and can then be put in relations to objects in the world; however, meanings are better conceived of as the cognitive precursors of the acts by means of which extensions are established, meanings are in fact better understood as a special sort of "cognitive events" (to use a notion from Soames 2015). We should observe that Pietroski's proposal concerning sentence-meaning amounts to a radical re-interpretation of the notion of compositionality in current formal semantics approaches. This re-interpretation is based in turn on a radical refusal of the cognitive relevance of higher types and on the claim

that the types with which language works are essentially predicative in nature. This position is reminiscent of the position endorsed by James Higginbotham in the 1980s, based on the insight that a Davidsonian event-semantics limited to the expressive tools of first-order logic was better suited to address the compositionality issues in language than the whole unconstrained apparatus of the lambda-calculus (see, for example, Higginbotham 1985).

In the recent philosophy of language and mind the debate between what is internal and what is environmental in the definition of semantic content is often formulated in the terms of the opposition between *broad* and *narrow* content. Broad content is content understood in such a way as to encompass both the environmental and internal factors that are relevant to it. Narrow content is content understood from a purely internalist perspective—what remains of broad content once stripped off of all of its environmental features. The debate concerns whether there is a meaningful distinction between broad and narrow content and, if there is such a distinction, how to best formulate the two notions. At the two extreme ends of this debate we find strong externalism and strong internalism. On the one hand, strong externalism entails that there is no such thing as narrow content for all content ultimately depends on environmental factors. According to this view, actual Kazimir and Twin Kazimir simply express distinct broad contents when uttering the word “water”. Conversely, strong semantic internalism entails that there is only narrow content. According to this view, actual Kazimir and Twin Kazimir express the exact same meaning when uttering the word “water” because they are characterized by identical internal states. The intermediate position consists in maintaining that there is a distinction between broad and narrow content. When actual Kazimir and Twin Kazimir utter the word “water” they express different broad contents, as they are referring to different objects in their environment, but also share a common narrow content, as their use of the word “water” is a function of their common internal state. The challenge for those who adhere to this position is how to define narrow content. Fodor has been amongst the most eminent proposers of a functional analysis of narrow contents whereby narrow contents are like Kaplanian characters, mapping environments into broad contents (see Fodor 1987). As we saw, this view has been criticized by Stalnaker and Block (see Stalnaker 1989 and Block 1991). Other approaches include the analysis of narrow contents in terms of *conceptual roles* (see Block 1986) and in terms of *epistemic possibilities* (see Chalmers 1996, 2002, 2003).

## References

- Block, N. (1986). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy*, 10, 615–678.
- Block, N. (1991). What narrow content is not. In B. Loewer & G. Rey (Eds.), *Meaning in mind: Fodor and his critics*. Oxford: Blackwell.
- Chalmers, D. (1996). *The conscious mind*. New York/Oxford: Oxford University Press.
- Chalmers, D. (2002). The components of content. In D. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings*. Oxford: Oxford University Press.

- Chalmers, D. (2003). The nature of narrow content. *Philosophical Issues*, 13, 46–66.
- Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
- Fodor, J. A. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.
- Higginbotham, J. (1985). On semantics. *Linguistics Inquiry*, 16, 547–593.
- Pietroski, P. M. (2005). Meaning before truth. In G. Preyer & G. Peter (Eds.), *Contextualism in philosophy* (pp. 253–299). Oxford: Oxford University Press.
- Pietroski, P. M. (2018). *Conjoining meanings. Semantics without truth-values*. Oxford: Oxford University Press.
- Ramchand, G. (2016). *Situations and syntactic structures*. Cambridge, MA: MIT Press.
- Soames, S. (2015). *Rethinking language, mind and meaning*. Princeton: Princeton University Press.
- Stalnaker, R. C. (1989). On what's in the head. *Philosophical Perspectives*, 3, 287–316.

**Part III**  
**Meaning and Perception**

# Chapter 31

## Beyond the Cartesian Divide



### Meaning as Interface

Janus is among the oldest divinities venerated in ancient Rome. He is also amongst the most original, as he is one of the few Roman gods that was not imported from the Greek pantheon but descends from an autochthonous tradition. Janus is canonically regarded as the god of transitions. In fact, he is typically depicted as having two faces oriented towards opposite directions. To this feature he owes his famous epithet of *bifrons*, literally meaning “two-faced”. Because of his capacity to look, at once, inward and outward, his statue was placed on top of doors, gates, bridges, and other passages. He also presided to the gates of his temple in Rome, which were kept open whenever Rome was at war and closed during times of peace. Janus is the god of transitions also in a more abstract sense. As he can connect not only inside and outside but also past and future, he is the god of endings and beginnings. To him, Romans dedicated the morning—the beginning of the day—as well as the first day of every new year and month. According to some, in fact, it is to him that we owe the name of the first month of the year, January. It is also to him that Romans offered their prayers whenever starting a new enterprise, a journey, or a ritual.

In part II, we investigated the metaphysical foundations of linguistic meaning by considering and comparing two alternative views—semantic externalism and semantic internalism. The first view claims that meaning is grounded in the external world of natural things. The second claims that meaning is grounded in the mind of speakers. Both views come with strengths and weaknesses. Semantic externalism succeeds in capturing the material connection between meaning and the external environment in which it occurs but fails to account for the cognitive relation speakers hold to it—how they learn it, entertain it, and how they use it to express their inner feelings, emotions, and sensations. Semantic internalism, conversely, manages to explain how speakers relate cognitively to meaning but fails to capture its

relation to the external world. As we saw by the end of part II, a bare sum of the two views is also insufficient.

In the third and last part of our book, we will make an attempt at overcoming this impasse by exploring an approach that is alternative to both the hypothesis of semantic externalism and that of semantic internalism. The hypothesis we will try to support is that the metaphysical foundations of linguistic meaning are to be found neither in the external world of natural things nor in the inner realm of the mind but at the *interface* between the two. In a way, the view that we will discuss regards meaning as a sort of Janus *bifrons*—a two-faced entity that is able to connect, at once, the external world of natural things with the inner realm of the mind. In order to achieve our goal, however, we will first need to undertake a radical revision of the traditional Cartesian understanding of the divide between world and mind.

## World, Mind, and Perception

What makes world and mind so difficult to reconcile with one another? To a significant extent, the source of the trouble resides in an incorrect view of *perception*. This is a view that has a long tradition in the history of western thought. It can be traced back to Aristotle although its most eminent modern representative is Descartes. For reasons that will be made clear in the following chapter, we will refer to this view as the *image view* of perception.

As we will see in the course of this part of the book, the image view of perception is not only highly problematic but also factually wrong. The contemporary understanding of perception—the result of a number of empirical and theoretical advances in neurobiology and the cognitive sciences—offers us a very different picture of how perception actually works. Perception, scientists tell us today, is not what divides world and mind but, in fact, the glue that holds them together.

A closer investigation of how perception really works also shows that the basic logical and metaphysical ingredients of perception are remarkably similar to those we find in natural language meaning—as we have identified them in the course of the previous two parts of the book. This observation offers us a formidable solution to the problem of how to reconcile world and mind in linguistic meaning. Meaning, we will claim, is grounded neither in the external world of material things nor in the inner world of the thinking mind. Its foundations are rather to be found at the point of juncture of these two realms—that is, in perception.

The following chapters are organized as follows. In Chap. 32, we discuss the traditional view of perception along with its shortcomings. In Chap. 33, we introduce the modern scientific view of perception and discuss its main conceptual consequences. In Chaps. 34 and 35, we show how the logical and metaphysical framework at the basis of sensory perception can be applied to natural language meaning and how doing so allows us to reconcile the material and cognitive dimensions of linguistic meaning. We draw some general conclusions in Chap. 35.

## Chapter 32

# World, Mind, and Perception (The Incorrect View)



### The Image View of Perception

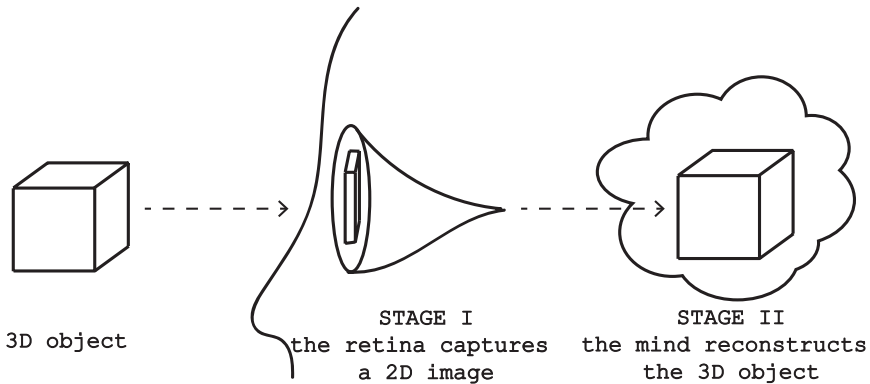
The image view of perception is the most traditional view of perception. According to this view, perception is the product of the workings of two fundamentally different and largely independent systems. The first is the *sensory system*, made of receptors—such as eyes and ears—whose goal is to mechanically record external energy patterns—such as light or sound waves. The second is the *conceptual system*, whose goal is to interpret the passive record provided by the sensory system by classifying its content around *concepts*—classificatory notions such as shape, color, distance, pitch, intensity and so on.

Let us consider a couple of examples from the domain of vision. Consider the case in which an observer is looking at a three-dimensional object, such as a cube (Fig. 32.1). According to the image view of perception, the process of perceiving the three-dimensional cube proceeds in two stages. First, the two eyes record the distribution of light that hits them and each eye produces a flat, two-dimensional image. Then, the mind reconstructs the three-dimensions of the object by contrasting the two two-dimensional images provided by the eyes (a process called *stereopsis*).

As a second illustrative example, consider the capacity to distinguish an object from its background. Also in this case, the process proceeds in two steps. First, the retinas record a two-dimensional image, where object and background are squeezed on a single flat surface. Then, the mind reinterprets the flat image as the product of two separate elements, the object and its background.

Both these examples illustrate Descartes' idea that perception is the product of the sequential workings of two independent processes performed by two independent systems. The first is the process of recording external information, which is performed by the sensory system. The second is the process of interpretation of such record, which is performed by the conceptual system—the mind.





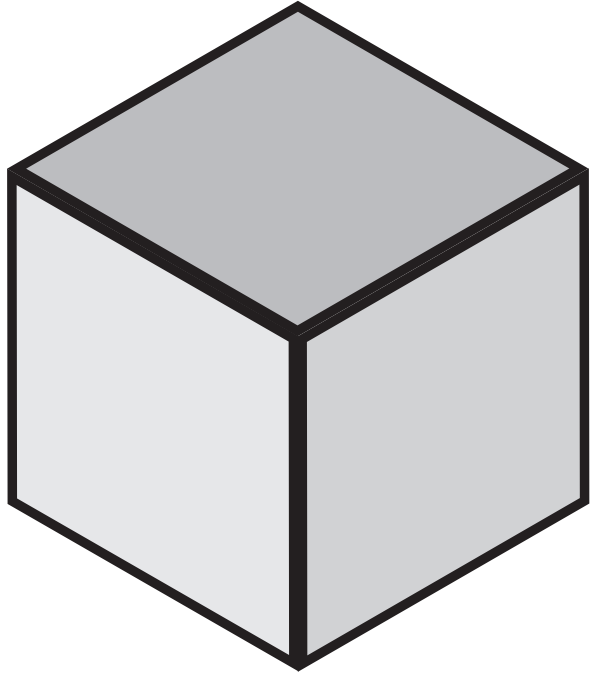
**Fig. 32.1** A depiction of how an observer perceives a three-dimensional object according to the image view of perception. The process proceeds in two steps. In the first step, the retinas in the eyes record the distribution of light that hits them and each eye produces a flat, two-dimensional image. In the second step, the mind reconstructs the three-dimensions of the object by contrasting the two two-dimensional images provided by the eyes

According to the image view of perception, the two processes are essentially different: The first is strictly *material*; the second is strictly *mental*. Sensory systems, on the one hand, are mere devices of *transduction*. They record passively, mechanically, and automatically the pattern of energy that is incident upon them. The process of conceptualization, on the other hand, is genuinely mental. It requires a conscious interpretive effort on the part of the observer in organizing the raw products of the sensory systems around concepts.

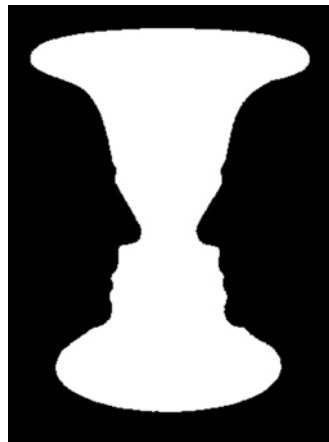
The motivation for distinguishing the process of sensory registration and that of conceptual interpretation is provided by Descartes' fundamental observation that perception can be deceptive. Consider Fig. 32.2. Is it the representation of a convex or concave cube? That is, are we observing the inner or outer faces of the cube? The answer is that it can actually be both, depending on how we conceptualize it (in fact, with some practice, it is easy to learn to decide at will whether to interpret the figure as an outward-looking or inward-looking cube). The image is, in other words, *ambiguous*. This ambiguity is naturally explained by the image view of perception as a product of the fact that the way we see things *as* is not only the product of the mechanical registration of a sensory input but also of an independent mental effort of conceptual interpretation.

Similar ambiguities occur when distinguishing a figure from its background. A famous example is that of the face-vase illusion, by Edgar Rubin (Fig. 32.3). Is this the image of a white vase against a black background or that of two black faces against a white background? Again, there is no right or wrong answer. Rather, the image is ambiguous and the interpretation we assign to it depends on how we conceptualize it. Illusions such as these show us that perception is the product of a genuinely mental effort of conceptual interpretation. What we perceive things as depends not only on the mechanical operations that are performed by the senses but also on the interpretative effort that is contributed by the mind.

**Fig. 32.2** An ambiguous figure, which can be interpreted either as a convex or concave cube



**Fig. 32.3** An example of Rubin’s illusion. The figure can be interpreted as representing either a white vase against a black background or two black faces against a white background



## Problems with the Image View of Perception

The image view of perception is extremely intuitive. For one thing, it fits squarely with our intuitions about perceptual illusions. In fact, it is the view maintained—more or less explicitly—by a number of modern and contemporary scholars. It is at the foundations of Russell's distinction, which we reviewed in Chap. 24, between acquaintance—the mechanical process of reacting indexically to external events—and description—the cognitive process of classifying the information detected on the basis of conceptual features. It is also the view maintained in more recent philosophical works such as John McDowell's (1994) book *Mind and World*. Yet, as intuitive as it may appear at first, it also suffers from some profound problems. In this section, we will review two of the most important.

The first problem is epistemic. As we saw above, according to the image view of perception, what we perceive things as is the product of two distinct ingredients: the raw sensory data collected by the senses and the mental process of interpretation performed by the conceptual system. Hence, what we perceive things *as* is never solely determined by what things *are*, because perception is also the product of an independent mental effort. This seems correct, as demonstrated by the occurrence of perceptual illusions. This also means, however, that we can never reconstruct what something is from what we perceive it as. After all, as proven by the case of perceptual illusions, we can always be wrong in how we conceptualize the raw sense data. If correct, then, the image view of perception entails that perception is an unreliable source of information about the external world. There simply is no principled way to prove that what we perceive things as corresponds to what things really are because the way we perceive thing as is inevitably vitiated by a component of independent conceptual interpretation. Ultimately, our senses cannot be trusted. As is well known, Descartes regarded this conclusion as astounding but not necessarily problematic. For him it was, rather, a reality to be confronted with. Ultimately, the only thing we can know with certainty is our own mind and it is upon the foundations of such realization that we must reconstruct the edifice of human knowledge. Yet, even if we leave Descartes' philosophical preoccupations aside, we must at least grant that the conclusion that perception is an unreliable source of information about the world contrasts dramatically with the simple observation that perception is used by humans—as well as a vast variety of other organisms—to navigate and exploit their environment in a number of different ways. The fact that humans (like many of their precursors) have managed to survive and evolve in their environmental niche for a few million years suggests that perception must be able to represent reality correctly at least to the extent that it allows humans to effectively cope with it. How would this be possible if perception was irremediably detached from reality, as the image view of perception would have us conclude?

The second problem for the image view of perception is a genetic problem. It has to do with the source of the mental concepts that allegedly allow us to interpret the raw products of sensory systems. Where do these concepts come from? The question can be understood in both a phylogenetic and an ontogenetic sense.

Phylogenetically, the question is how concepts have emerged in the human being in the course of its evolution. Ontogenetically, the question is how concepts emerge in the course of the development of individual human beings. Both questions are extremely problematic for the advocates of the image view of perception because, according to such view, concepts cannot come from perception itself. Concepts, in fact, are a precondition for perception to take place. If there are no concepts, there is no perception. To be able to perceive an object as a black square, one must already possess the concepts of “black” and “square”. A creature without these concepts simply lacks the conceptual tools to interpret what it sees as the representation of something that is black and that is a square. But how can one acquire the concepts of “black” and “square” if not by observing them in the real world in which they occur? As we all know, Descartes had an answer to this question as well. Concepts are divine. They are instilled by God in the human soul. In fact, he believed that animals do not have them. As animals do not have a soul, so they lack the ability to contemplate God and, with it, the conceptual apparatus only God can grant. Independently of one’s inclinations concerning God, the natural world, and its living creatures, it is difficult to accept Descartes’ solution as one that has any empirical value by today’s scientific standards. What would be preferable is an explanation of the human ability to perceive the world as organized around some descriptive categories rather than others as a function, among other things, of the way humans interact with their environment.

Both of these problems are topics of important philosophical discussions and a vast literature. From our perspective, they help us demonstrating how the image view of perception traces an irreparable hiatus between world and mind. On the one hand, the world is bound to remain inexorably inaccessible to the mind, for what we perceive the world as is not what the world really is. On the other hand, the mind is bound to remain a divine realm of immaterial concepts whose origin cannot be traced back to the natural world itself.

## References and Remarks

For an introduction to Descartes’ views on perception in the light of the traditions that preceded and followed his work we recommend Chap. 2, sections I and II, of Mohan Matthen’s essay *Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception* (Matthen 2005).

## References

- Matthen, M. (2005). *Seeing, doing, and knowing. A philosophical theory of sense perception.* Oxford: Oxford University Press.
- McDowell, J. (1994). *Mind and world.* Cambridge, MA: Harvard University Press.

# Chapter 33

## World, Mind, and Perception (The Correct View)



The image view of perception is not only problematic for the reasons we have mentioned in the previous chapter. It is also factually wrong. As we will see in this chapter, the modern scientific study of perception provides us with a very different picture. Sensory systems are not the passive receptors of external inputs Descartes thought them to be. They are, in and by themselves, instruments of categorization and interpretation, which have emerged in the course of evolution with the specific purpose of providing humans—in the same way as other living organisms—with useful information about their environment, serving them in the challenging task of surviving and flourishing in their environmental niche.

### Perceiving Features

Contrary to what the image view of perception maintains, sensory systems are not transducers. Their function is not that of converting patterns of physical energy—as they are provided by the external world—into states of the brain—as they are then interpreted by the mind. Rather, sensory systems are *filters*. Their explicit task is that of selecting information from the external world by testing specific hypotheses.

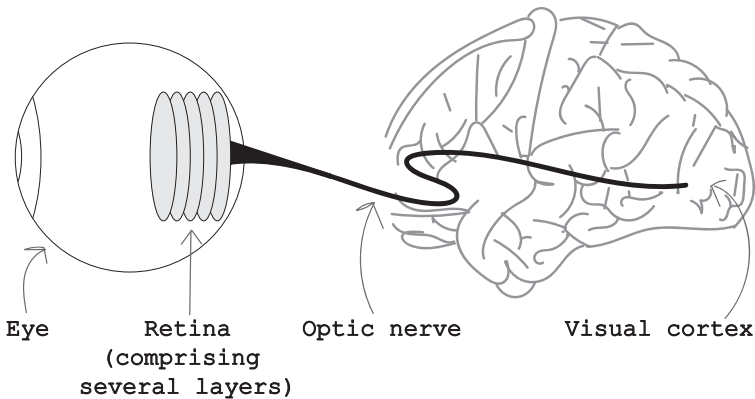
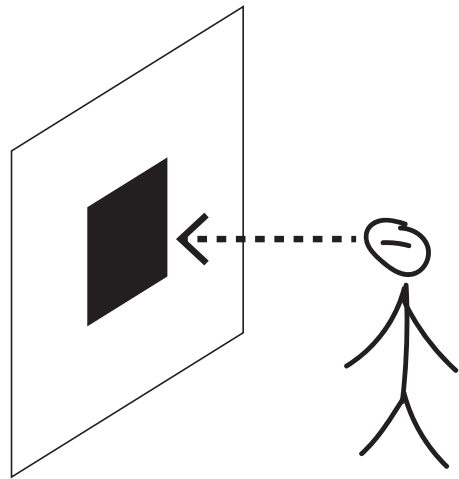
In order to provide a concrete example of the workings of an actual sensory system, we will consider, in what follows, a simplified model of the visual sensory system as it is found in humans. By doing so, we will be able to highlight its essential logic, which is the same we find in other sensory systems—such as hearing, touch, taste, and smell—in humans as well as in other creatures.

Let us consider a human being, say, Kazimir, visually attending at a black square on a white background (Fig. 33.1). The process that leads Kazimir to perceiving the black square begins with the light hitting Kazimir's eyes and, more precisely, his retinas—the light-sensitive coats of receptive tissue within his eyes. Each retina comprises a number of adjacent layers. Immediately after the light has impacted the

first layer of the retina, the information captured by the cells on the retina is transmitted to the neurons in the subsequent layers in the retina. These are the first links in the downstream chain of synaptic connections that bring visual information from the retina to the optic nerve and, then, to the visual cortex in the brain (Fig. 33.2). Each neuron in each layer of the downstream flow of synaptic connections receives its input from a group of cells in the upstream layer. The group of upstream cells from which a downstream neuron receives its input is called the neuron's *receptive field* (Fig. 33.3).

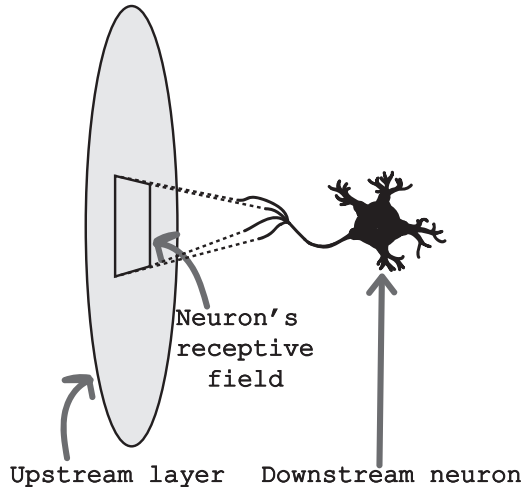
How is visual information transmitted from the first layer of receptors in the retina to the next downstream layer of neurons? Before looking at the experimental findings, it is useful to consider how the image view of perception would predict this process to take place. If the goal of the visual sensory system was that of merely

**Fig. 33.1** Kazimir visually attending at a black square on a white background



**Fig. 33.2** The downstream chain of synaptic connections that bring visual information from the retina to the optic nerve and, then, to the visual cortex in the brain

**Fig. 33.3** A retinal layer and a neuron in its immediate downstream. The neuron's receptive field is the portion of the retina from which the neuron receives its input



registering external information, we would expect the neurons in the downstream layers to react proportionately to the energy of the light that hits their receptive field. If, for example, a receptive field was hit by an intense amount of light, we would expect its corresponding neuron to fire more strongly, indicating a powerful source of light or a highly reflective surface. If, conversely, the receptive field was hit by a weak stream of light, we would expect the neuron's response to be weak, indicating a poor source of light or a highly absorbing surface.

We cannot blame Descartes for not having put more effort into testing his views experimentally. Not only did he lack the technology to do so, he also lacked the very notion of a neuron. These notions and instruments, however, were available to neuroscientists such as Horace Barlow at Cambridge University and Stephen Kuffler, David Hubel, and Thorsten Wiesel at Harvard. In the 1950s and 1960s, they were among the first to use the then emergent technology of single neuron recording to test experimentally the functioning of the neurons in the visual sensory system. Their results set the foundations for the computational models of visual perception that would then be developed in the following decades.

The picture that emerges from the experimental observation of neurons in the visual sensory system is very different from the one predicted by the image view of perception. Neurons in the visual sensory system do not act as registers of information, reacting proportionally to the amount of energy captured by their receptive field. Rather, neurons are best described as computing devices whose function is to test whether the information provided by their corresponding receptive field does or does not meet a certain condition. As an example, let us consider a neuron whose function is to detect the color black in its receptive field. This neuron operates as follows: It fires an electric charge if there is black in its receptive field, but remains still otherwise. We can describe the function performed by this neuron as that of testing the following condition in its receptive field:

## (B) There is black

If there is black within the neuron's receptive field, then condition (B) is met and the neuron fires a synaptic signal to the next layer of cells. If, conversely, there is no black within the neuron's receptive field, then condition (B) is not met and the neuron does not fire a synaptic signal. As a further example, consider a neuron whose function is to fire if, and only if, there is a horizontal edge in its receptive field. We can describe its function as that of testing the following condition in its receptive field:

## (H) There is a horizontal edge

If the neuron detects a horizontal edge in its receptive field, then condition (H) is met and the neuron fires. If there is no horizontal edge, then condition (H) is not met and the neuron does not fire.

As we see from these examples, the output of a neuron in the visual system is not proportional to its input, as the image view of perception would predict. It rather indicates whether its input—the neuron's receptive field—meets or fails to meet a certain condition. Abstractly, we can describe the function performed by such a neuron as that of a function from state of affairs into truth-values. The input state of affairs is the receptive field. The output is the response—the firing or not firing of an electric charge to indicate whether the input state of affairs does or does not meet the condition at hand. A useful metaphor, due to Mohan Matthen, is that of the lookout. Standing on top of the ship's main mast, his task is to shout "Ship ahoy!" whenever he detects a ship on the horizon. Similarly to the lookout, a neuron in the visual domain is charged with the task of shouting—by firing an electric charge to the next layer in the downward stream—whenever it detects a certain condition in its receptive field.

We see that the function of the visual sensory system is not that of providing a faithful translation of the input, as the traditional view of perception maintained. Even its more peripheral receptors are involved, instead, in a process of classification of the input around specific conditions. This observation is true of the visual sensory system as well as of the other sensory systems. Indeed, this finding puts us in front of a new question. What determines the conditions that are relevant to our sensory systems? According to current estimates, the human retina alone hosts as many as fifty-five different types of neurons testing as many conditions. Yet, these conditions are but a tiny subset of the vast number of physical features that could potentially be tested in our material environment. What determined such subset? The answer to this question is: *nature*. Our sensory systems have developed in the course of our evolutionary history with the explicit function of helping us survive and, possibly, flourish in our environmental niche. Accordingly, they have developed to collect from the environment the information that helps us achieve these goals and discard the information that does not.



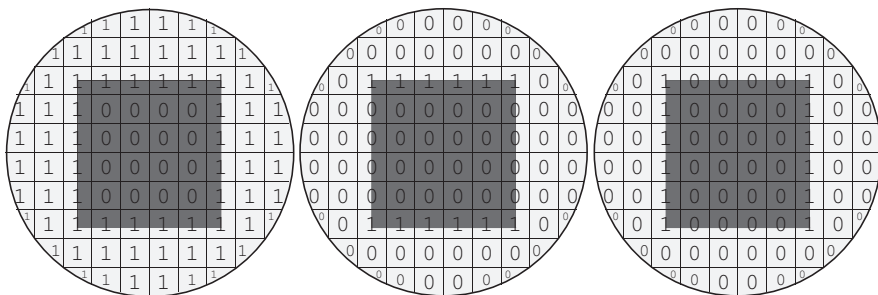


As we said, different layers test different conditions and, therefore, deliver different feature maps. Let us consider, for completeness' sake, the other feature maps that contribute to Kazimir's visual perception of the black square. In the simple scenario we are considering, at least three other conditions, besides (B), are relevant to Kazimir's perception of the black square. These are condition (W), that there is white, condition (H), that there is a horizontal hedge, and condition (V), that there is a vertical edge.

- (W) There is white
- (H) There is a horizontal edge
- (V) There is a vertical edge

Each of these conditions is tested by a corresponding layer of neurons which, in turn, provides a corresponding feature map (Fig. 33.5). The leftmost disk represents the feature map testing feature (W), which is characterized by the positive value 1 whenever a receptive field contains an instance of the color white. The disk in the center reports the feature map testing condition (H), which is true of all receptive fields that contain an instance of a horizontal edge. The rightmost disk reports the feature maps testing condition (V), which is true of all receptive fields that contain an instance of a vertical edge. The outcome of the different layers of neurons in the retina is, finally, a series of feature maps each reporting on the testing of a different condition. These feature maps constitute the information that is transmitted through the optic nerve to the visual cortex.

The finding that the retina delivers a set of distinct feature maps has traditionally posed an extraordinary challenge to scientists. When Kazimir looks at the black square, what he sees is not a set of distinct features but, in fact, a single object, comprising all the different features at once. How does the visual sensory system manage to combine all these different features into a single object of perception? This problem is known among specialists as the *binding problem*—the problem of describing how the visual sensory system manages to integrate the distinct feature maps it produces to deliver a single perceptual experience. The binding problem is made exceptionally difficult by the observation that the different feature maps



**Fig. 33.5** The feature maps provided by the layer of neurons testing, from left to right, conditions (W), (H), and (V)

provided by the sensory system are strictly dissociated from one another. This has been proven experimentally by showing that the effects of a single feature map can be blocked selectively. Such cases are either the result of a specific pathological condition affecting the ability to consciously access a specific feature map or obtained in the lab by temporarily inhibiting—through artificial techniques—the neurons that belong to a specific layer. As a concrete example, let us imagine that we were to block the layer of neuron's in Kazimir's retinas that are in charge of testing condition (B). As a result, Kazimir would still perceive a square, but would fail to recognize it as black. Studies of this sort have proven that the different layers operate independently of one another and that the feature maps they produce are strictly dissociated from one another. When addressing the binding problem, we should also not forget that feature maps are not in and by themselves images. Even though we represented them visually as grids, distributed on a two-dimensional space, they are, in reality, clusters of synaptic activations, more akin to data sets—lists of 0's and 1's—than to two-dimensional images. The solution to the binding problem, hence, cannot be a procedure that merely collapses the different layers one on top of the other into a single final image. Against this type of solution, multiple experiments have also shown that feature maps are, more often than not, characterized by different resolutions.

The challenge posed by the binding problem is, henceforth, that of explaining how the visual sensory system manages to combine the distinct data sets provided by the distinct feature maps into a single integrated visual experience. Notably, the binding problem emerges also with the other sensory systems, besides vision. An illustrative example is that of hearing. Similarly to the system of visual perception, also the system of auditory perception analyzes its input by producing different feature maps, each encoding a different feature of the input—such as pitch, intensity, and so forth. Yet, what one hears is not a set of distinct features, but a single sound comprising all these different features at once. The binding problem, in this case, is how the auditory sensory system manages to produce the experience of a single sound starting from these dissociated feature maps.

## Perceiving Objects

Sensory systems allow us to perceive the world around us as made of integrated *objects*. How do they manage to do so? Before addressing the question from the modern scientific perspective, it is informative to consider how the traditional view would go about it. Once again, let us use visual perception as our paradigmatic case. In the framework of the image view of perception, objects are a derivative notion. According to the image view of perception, perception proceeds in two steps. In the first step, the visual sensors translate the input into a flat, two-dimensional representation. Then, in the second step, this representation is interpreted by the conscious efforts of the conceptual mind. The notion of an object belongs to this second stage. It is only upon conceptual reflection on the properties of the flat image that the

observer is able to single out the objects that occur in it. Consider again, as an example, Kazimir observing the black square on the white background. According to the image view of perception, Kazimir is first provided by his sensors with a flat image, made of black and white areas. Then, by conscious reflection, Kazimir is able to single out the black areas in the flat image as belonging to a single object—the black square—that is distinct from its background. As we see, according to the image view of perception, in order to identify something as an object, we must first conceptualize its properties.

There are at least two reasons why this is incorrect. First, we know now that the visual sensory system does not produce a flat reproduction of the input. As we saw in the previous section, the main task of the peripheral visual sensory system is that of producing feature maps—independent data sets, each reporting on the testing of a different condition. These feature maps do not offer, in and by themselves, sufficient information to support the singling out of individual objects. On the contrary, they raise the problem of how they are integrated with one another to produce the image as it is perceived by the observer—what we referred to above as the binding problem. The second reason is that the notion of object cannot be ancillary to that of property: the world we live in is a dynamic one and so are its objects. Things change their properties over time, sometimes very rapidly, and, yet, we recognize them as being the same object. They may change shape, color, and location. Sometimes they may disappear from our visual field and, then, suddenly reappear. Sometimes they may be only partially visible. Yet, as observers, we are extremely good at recognizing them as integrated objects. How do we manage that?

The ability to recognize objects across time, despite their changes in location, shape, color, and other properties, is referred to by specialists in the field of perception as *tracking*. Tracking has been studied in great detail in recent decades by means of a number of sophisticated experimental paradigms, in good part due to the cognitive scientist Zenon Pylyshyn and his team. To introduce the notion of tracking, we will rely on an example we are all familiar with: motion pictures.

We are all familiar with movies and animations and we all know that the effect of motion we observe when watching them is, in fact, illusory. In reality, what we are witnessing is a sequence of still pictures. Yet, because of the high speed at which the pictures in the sequence are presented, we are tricked into regarding them as the dynamic depiction of a single, uniform motion. In the technical jargon of motion pictures, the still pictures that compose a movie or animation are called *frames* and the speed at which they are projected is called *frame rate*. Nowadays, the most commonly adopted frame rate in the production of movies and animations is of 24 frames/s. At this rate, the visual sensory system is unable to recognize the distinct frames in the sequence. All it can perceive is a smooth, uniform motion. In more recent times, producers have experimented with higher frame rates—of, for example, 30, 48, or even 60 frames/s—in an attempt to achieve an even smoother feeling of uninterrupted motion. However, what is truly fascinating about the mechanics of motion pictures is, quite to the contrary, that it works even at extremely low frame rates—such as of 8, 6, or even 5 frames/s. These low frame rates are used rarely and almost exclusively in the production of hand-crafted animations.

The main reason for adopting them is practical: At such low speeds, the artist can cover longer stretches of time by drawing fewer frames. More often than not, however, there is also an artistic intention behind their use. They allow the magic of animation to transpire more vividly. When we watch an animation at a very low frame rate, it is easy, if not unavoidable, to become aware of the distinct frames following one another in the sequence. Yet, we simply cannot stop our eyes from perceiving the sequence as a motion, no matter how clumsy and stuttering the slow sequence of frames is. To many masters of the field, this is the true beauty of animation. As it takes so little to disclose its effect, it can be made with the humblest means—sometimes only a pencil and a stack of papers.

The cognitive skill that is at work when we watch a movie or animation is the one we referred to above as tracking—the ability to follow an object across time and space, despite its changes in properties. Tracking is characterized by three main features. The first is that it is performed by our visual sensory system automatically and unconsciously. This has been proven by Pylyshyn and his colleagues through a number of sophisticated experiments. We can appreciate it also in the context of our example. As we saw a moment ago, we may very well put all of our conscious efforts into conceptualizing the slow, stuttering animation we are watching as a sequence of static pictures. None of these efforts will suffice to stop our visual system from seeing it as a motion. The second property of tracking is that it can be easily performed with more than one object at once. We experience this every time we watch a movie or animation representing more than one moving object. In Pylyshyn's experiments, this feature has been highlighted by testing observers in tasks where they have to keep track of several objects on a screen. Even in extremely challenging settings, where the objects are all rapidly moving, transforming, and sometimes even disappearing behind occluding surfaces, participants not only perform very well, but also report that the task is relatively easy to perform. The third property of tracking is that it does not rely in any way on the ability of the observer to represent the properties of the objects tracked. The fact that an object may change its shape, color, location, or any other of its characterizing features does not impede in any way the ability to track it. In fact, the most substantial proof for this claim comes from the finding that the tracking of an object can be performed effortlessly at a speed that is significantly higher than that which would be required by the visual sensory system to detect any of the object's properties. We are, in other words, quicker and better at telling that something is an object than we are at identifying its properties.

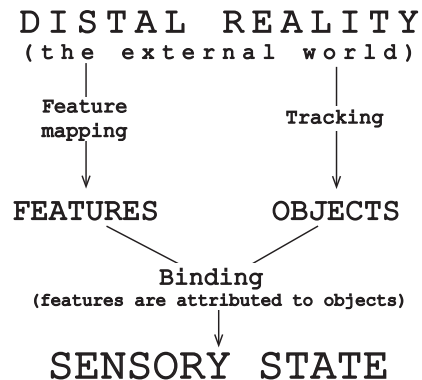
The image view of perception, as we saw above, maintains that the notion of an object of perception is a derivative one. We can recognize something as an object only once we conceptualize its properties. The study of tracking proves the exact opposite. The ability to identify objects in our environment is even more primitive than that of detecting their properties. According to experts today, in fact, object tracking is the most primitive contact we entertain with the world around us. A parallel that is often used to explain the notion of tracking is one that has to do with language. The relation of tracking, it is claimed, is like an *indexical* relation. As we saw in Chaps. 21, 22, and 23, indexical relations are important in language because

they determine the connection between an expression and its environment of use. An indexical relation is what connects the expression “I” to whomever happens to be the person speaking in the context in which “I” is uttered. Crucially, as we observed in Chap. 21, indexical relations are not the same as descriptions. As proven by Kaplan, the link that connects an indexical expression to its reference is direct and unmediated. “I”, for example, refers to the person speaking but does so regardless of the fact that this person is the person speaking. The direct, unmediated link that we find in indexical reference is the same link that we find in tracking. Similarly to natural language, our sensory systems have a way of tracking objects in their environment that operates irrespectively of the properties of such objects.

## Perceiving Propositions

Tracking also provides a solution to the binding problem. So far, we have seen that the visual sensory system starts off by collecting two distinct types of information, objects and properties. Objects, as we just saw, are detected through the mechanism of tracking. Properties are detected through the mechanism of feature mapping. The two mechanisms—object tracking and feature mapping—operate in complete independence from one another. Their dissociation has been proven by various experiments and also by the observation that they activate distinct neural pathways in the brain. Once objects have been tracked and properties have been mapped, the next step in the downstream flow of information that produces the final perceptual experience consists in binding them together. This step consists in *attributing* the properties detected by the feature mapping mechanism to the objects detected by the tracking mechanism. We will refer to the product of this operation as a *sensory state*—a description of a set of objects and their properties. The procedure is exemplified in Fig. 33.6.

**Fig. 33.6** Sensory systems obtain information about the external world by means of the two mechanism of feature mapping and object tracking. Features and objects are then bound together to produce sensory states



To exemplify this process, let us return to the scenario where Kazimir is observing a black square on a white background. Kazimir's visual experience starts off, as we saw, by collecting two types of information. The tracking mechanism detects the presence of an object in Kazimir's visual field. The feature mapping mechanism produces a number of feature maps. Four such maps are relevant to the current example—those produced by testing conditions (B), (W), (H), and (V). The following step—the one we are considering now—consists in combining these elements together by attributing the four features mapped to the tracked object. This process produces a sensory state describing an object characterized by the following attributes: vertical edges on its right and left, horizontal edges above and below, black inside its edges, and white outside of them.

We can better appreciate the notion of a sensory state by using, once again, a parallel with language. Sensory states are equivalent to *propositions* and, more exactly, *de re* propositions. A proposition, as we know, is the description of a state of affairs. A *de re* proposition is the description of a state of affairs where a property is attributed to a *res*—that is, an object. The *de re* proposition expressed by the sentence “Kazimir is a painter”, for example, describes a state of affairs where the property of being a painter is attributed to the *res* Kazimir. In a parallel fashion, sensory states describe states of affairs where one or more properties are attributed to one or more objects of perception. For example, the sensory state produced by Kazimir's visual sensory system in the example we are considering describes a state of affairs where the object in Kazimir's visual field is attributed the properties of being a square, being black, and being surrounded by white.

Like propositions, sensory states are *representational* and *truth-conditional*. They provide a description of things as being such and such and their actual truth-value depends on whether the things they describe are actually such and such. Kazimir's sensory state, for example, provides him with a representation of the visual field in front of him as being populated by a black square on a white background. Whether such representation is correct depends on whether there is a black square on a white background in the actual circumstances captured by his visual field.

The parallel with language is correct in yet another way. As we know from part I, the *de re* propositions expressed by declarative sentences are the product of a *compositional* procedure. A *res* and a property are combined together, through the principle of *functional application*, to deliver a *de re* proposition. We find now that sensory states are the products of an equivalent procedure. As the principle of functional application combines objects and properties to deliver propositions, so sensory systems combine objects and properties to deliver sensory states.

It is the compositional nature of sensory states that help us address the binding problem. The distinct features that are produced by the distinct layers by testing distinct conditions are ultimately perceived as belonging to a single, integrated visual experience because they are attributed to one and the same object. It is as a result of this process that Kazimir's perceives the black square as a single bundle of different properties.

Compositionality also helps us explain the logical relations between different sensory states—the fact, for example, that we are able to judge that a black square



has the same color of a black circle and that a black square has the same shape of a red square. We are able to do so because different sensory states can be the product of the binding of equivalent properties—such as being black or being a square.

## Veridicality vs. Narcissism

As we know from the previous chapter, the fact that the products of perception are representations endowed with truth-conditions has been traditionally regarded as an indication that perception should not be trusted as a reliable source of information about the external world. To say that the products of perception have truth-conditions is to say that they can be false in the way they represent what they represent. That this is the case is demonstrated by the possibility of perceptual illusions—cases in which perception informs us that things are a certain way when, in fact, they are not. But then, how can we trust perception as a reliable source of information about the world if perception can be false in the way it represents the world? Was Descartes right in his skeptical stance?

The answer to these questions is that Descartes' skepticism is valuable only for as long as we maintain his view of perception. As we know, Descartes believed that perception proceeds in two steps. First, the senses provide an inert register of the external world. Then, this register is interpreted by the conceptual mind. It is during the course of this second stage that the mind judges which properties are true of the image and, as a result, a judgment of it which we can understand in the logical form of propositions. Propositions are, according to this framework, the final product of perception—what is obtained once the conceptual mind has performed its task of judging the inert register produced by the senses. Indeed, within this framework, the observation that the products of perception may be false of what they are meant to represent inevitably jeopardizes the whole enterprise.

We now know that this is not the way perception works. Sensory states have indeed the logical form of propositions. However, they are not the final product of sensory systems. They are, rather, instruments of data collection. Sensory systems, that is, do not collect information in order to later classify it around conceptual categories. Classification is, rather, the way in which sensory systems collect information in the first place.

The ultimate goal of sensory systems, in fact, is not that of providing *true* information about the external world. Their real goal is, rather, that of providing *useful* information about the external world—information, that is, that sustains the perceiver's survival in its environment. Truth and usefulness are very different notions. Truth is objective—the product of carefully excluding any source of subjective interference in the analysis of facts. Usefulness, in contrast, is explicitly subjective—selfishly dictated by the needs of whom it serves. As important as truth may be to scientists, historians, and philosophers such as Descartes, sensory systems have no interest in it. Their outlook on the external world is selfish and self-interested. The way they classify things is not dictated by a desire to faithfully



reproduce things objectively. It is rather dictated by the needs of the organism they serve. If, for example, Kazimir's visual sensory system classifies two different objects as both being black, that is not because the system wants to inform Kazimir that the two objects have comparable objective features. It is rather because the system wants to inform Kazimir that he can count the two objects as comparable from his own subjective perspective. As Matthen himself aptly puts it, "if a sensory system co-classifies two things, it is because these two things are comparable from the point of view of the organism they serve, not because anything else would so regard them, much less because *nature* would so regard them, whatever this might mean" (Matthen 2005, p. 60).

In a fitting parallel, the philosopher Kathleen Akins compares sensory systems to the *narcissist*—a person who can only see things from his own, self-interested perspective: "In a classic story, a narcissist goes to his therapist for his regular appointment. At the door he is met by another therapist in the same practice: there will not be a session that day, she informs him. The narcissist's therapist has been in a boating accident—she is alive, but in critical condition in the hospital. It is gently explained that there is good chance that the therapist will survive and recover, lead a reasonable normal life, perhaps even return to her practice. At this, the narcissist looks stricken, and wails: 'But why do these things always have to happen to ME?!?'" (Akins 1996, p. 345). As the narcissist, sensory systems are not interested in the question "what is it really like out there?" What is pressing to them is, rather, "how does this all relate to me?" The notions of truth and truth conditions are not the goal of sensory systems but only their means of data collection. Their ultimate goal is that of classifying environmental information in the way that best serves the selfish purposes of the perceiving subject.

## Conscious Perception and Sensorial Experience

The processes we have considered so far all happen unconsciously. The tracking of objects, the mapping of features, and their composition into sensory states are all performed by sensory systems without the intervention of any conscious effort. How are, then, the products of sensory systems presented to the conscious mind? How do we become aware of what we perceive?

Sensory states are presented to the conscious mind in the form of *sensorial experience*. To better understand this last step in the production of sensorial perception, let us return to the scenario where Kazimir is visually attending at the black square on the white background. So far, we have seen how Kazimir's visual sensory system constructs a sensory state—a *de re* proposition—describing an object endowed with the properties of being a square and being black. The last step in the downstream flow of information that leads to Kazimir to see the black square consists in translating this sensory state into Kazimir's sensorial experience of perceiving a black square. This experience is how Kazimir accesses, in his conscious mind, the information that he is visually attending at an object that is endowed with the properties

of being a square and being black. It is the *image* that is presented to Kazimir's conscious mind as a result of the unconscious workings of his visual sensory system. This information can, then, be used for further reasoning or be stored as a memory.

Sensorial experience is, therefore, what we perceive things *as*. As philosophers have long known, sensorial experience has rather peculiar properties. It usually referred to as a *phenomenal* object. The term, as it is used in this context, is due to Kant. In his Critique of Pure Reason, Kant famously distinguished between two types of objects: *noumena* and *phenomena*. Noumena are objects in and by themselves, as they exist in the real world of material things. Phenomena, conversely, are objects as they are experienced by the human mind. What motivates the distinction between noumena and phenomena is the observation that their relation is *arbitrary*. Take, as an example, colors. As the story goes, the way we perceive colors is determined by the wave length of the light that hits the retina. A long wave length of around 650 nm, for example, is perceived as red. A shorter wave length of about 500 nm is perceived as green. However, there is no principled reason why we should experience red the way we do, nor why we should experience green the way we do. Suppose that you were given two colored pencils—one red and one green—and that you were asked to draw a red circle and a green square on a sheet of paper. Envisage for a moment the drawing you would produce. Imagine now a person that, because of a specific congenital condition sees green whenever you see red and red whenever you see green. Suppose this person was also asked to draw a red circle and a green square. What would her drawing look like? It takes a moment of reflection to realize that, despite a common intuition for the opposite, her drawing would be identical to yours. The reason is that, when asked to pick up the red pencil, she would pick up the red pencil, no matter that the way red appears to her is different from the way it appears to you. Similarly, when asked to use the green pencil, she would use the green pencil, no matter that green looks to her the way red looks to you. Similar examples demonstrate that the way we experience the world is only arbitrarily related to the way the world is.

Of course, this fact was already clear to Descartes. In fact, it was on the basis of this observation that he motivated his skepticism for human sensorial experience. If the relation between the way we experience things (phenomena) and the way things are (noumena) is arbitrary, then the way we experience the world tells us nothing about what the world is. Indeed, it was this negative conclusion that caused the profound epistemic hiatus between world and mind we still experience in today's philosophical discourse.

We are now in the position to see that Descartes' negative conclusion is, once again, granted only within his understanding of perception. As we know, Descartes believed that our perception of the external world is produced in two steps. First, we produce, mechanically and unconsciously, a sensorial record of the external world. Then, we interpret this record through conscious effort, by organizing it around concepts. According to this view, our sensorial experience of the world precedes its interpretation. First, we have a sensorial experience and, then, we interpret it. We now know that this is not how perception works. The processes that produce our

experience of the external world are arranged, in fact, in the opposite order. Our perceptual systems first interpret external stimuli by classifying them into objects, properties, and, finally, sensory states. Then, this classification is presented to the observer in the form of a sensorial experience. Our sensorial experience of the world, hence, does not precede but, in fact, follows its interpretation. It is not where our sensorial understanding of the world begins but, rather, its final product. Consider again the case in which Kazimir is observing a black square next to a black circle. Kazimir does not classify the black square and the black circle as sharing a common property because, in his sensorial experience, they look the same. Rather, the black square and the black circle appear the same in Kazimir's sensorial experience because they have been classified by his visual sensory system as belonging to a common class.

The fact that sensorial experience stands in an arbitrary relation with the sensory states it conveys does not entail, henceforth, that it has no reliable informative value. To the contrary, the purpose of sensorial experience is precisely that of informing the conscious mind about sensory states. Once again, we can find a useful parallel with natural language. According to the view suggested by the scientific study of perception, a sensorial experience is like a word. As we saw as early as in Chap. 4, the relation between a word and its meaning is arbitrary. There is no principled reason why "table" means table and "chair" means chair. Yet, this does not entail that the words "table" and "chair" are meaningless or that we should not trust their meaning. It only shows us that their meaning is established by a *convention*. Similarly, there is no principled reason why we see red the way we do. This, however, does not entail that the way we experience the color red is meaningless. It simply demonstrates that the way we see red is a convention, adopted by our visual sensory system, to inform our conscious mind that the object we are observing belongs to a certain class.

The parallel between perception and language also helps us enlighten an important difference between the two domains. In language, the convention that connects a word to its meaning is the product of a stipulation made by the community of speakers. But what decides the relation between a sensorial experience and the sensory states it signifies? To whom do we owe the stipulation that a wave length of 650 nm is perceived as red whereas one of 500 nm is perceived as green? The answer to this question is, again: nature. The way we experience things sensorially is the result of how our sensory systems have developed, in the course of our evolutionary history, to provide us with useful information about our environmental niche and, in this way, help us survive.

In the end, perception does not constitute the unamendable epistemic divide between word and mind that Descartes had envisaged through his skeptical argument. To the contrary, the function of perception is explicitly epistemic. Its purpose is to inform the observer about the aspects of the external world that are useful to its survival. It is in order to achieve this goal that perception has developed throughout its evolutionary history. The contrast between warm colours, such as red, and cold colours, such as green, offers an informative example. It is true, as Descartes maintained, that there is no principled reason why red looks red and green looks green.

Yet, the distinction between the two has been functional to the evolution of humans—in fact, of primates in general—in helping them single out eatable fruits from the foliage that surrounds them.

## The Scientific Model of Perception

The model of sensorial perception that emerges from the discussion so far looks as in Fig. 33.7. The system proceeds through three main stages. In the first stage, it collects two sorts of information from its environment—objects and properties. Objects are detected through the mechanism of tracking, properties through the mechanism of feature mapping. In the second stage, properties are attributed to objects to produce sensory states—*de re* propositions that classify the objects in the environment on the basis of their properties. Finally, in the third stage, sensory states are signaled to the conscious mind of the perceiving cognitive agent in the form of sensorial experience.

The system is characterized by the following main features.

- (A) *Functional*. It exists in order to perform a precise function—that of supporting the survival of the perceiving organism providing it with useful information about its environment.

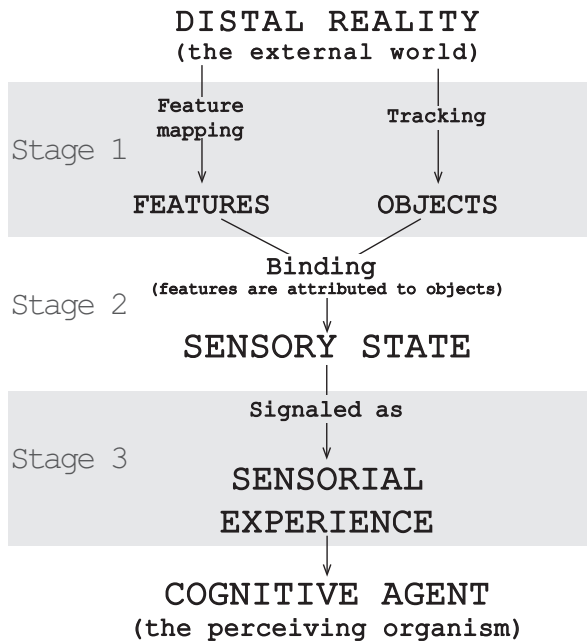


Fig. 33.7 The scientific model of sensorial perception

- (B) *Realist*. The system is resolutely anchored to the real world of material things in at least three respects. Firstly, the function of the system is to collect information about the natural environment around the organism it serves—the material reality that lies beyond its peripheral receptors. Secondly, the information collected by the system is explicitly addressed to the perceiver—a natural organism embedded in its natural environment. Thirdly, the system is anchored to the real world of natural things through its genesis. It has developed, in the course of its evolutionary history, in response to the problems the organism faced in its ecological niche. Ultimately, its genesis and function are products of the interaction between the ecology of the organism and that of its environment.
- (C) *Epistemic*. Besides being resolutely anchored to the external world, the goal of the sensory system is that of providing information to an epistemic agent—the organism it serves. Its products are, therefore, meaningful and informative to the agent that receives them.
- (D) *Propositional*. The information obtained by the system is packaged in the form of sensory states—de re propositions describing the environment in terms of objects and their properties.
- (E) *Compositional*. Sensory systems produce sensory states by binding together information of two primitive types, objects and properties. Sensory states are, hence, the product of a compositional procedure.
- (F) *Logical*. Because of their propositional and compositional nature, sensory states have logical properties and stand in logical relations to one another. It is in virtue of this properties and relations that perceivers are able to recognize a black square and a black circle as sharing a common property.
- (G) *Vivid*. The information gathered by sensory systems is made available to the subject's conscious mind in the form of a sensorial experience with phenomenal properties. Sensorial experience is the language sensory systems have developed in the course of their evolution to label sensorial information so that cognitive agents can access it consciously and memorize it for future use.
- (H) *Active*. Perception is neither a passive record of external events nor an exercise of the intellect. It is an action we perform. Sensory systems dynamically engage in the act of selection and classification of environmental information for the sake of supporting the organism in its survival.

All in all, the picture we obtain from the scientific study of perception is quite different from the one we inherited from Descartes. Perception is not mere sensory registration but the capacity to represent reality meaningfully. To fulfill its purpose, perception is endowed with a logical structure significantly richer than that of bare transducers, including high-level notions such as objects, properties, and propositions. At the same time, the fact that perception is representational is no more regarded as incompatible with a naturalistic view of the mind. Rather, the mind and its representational capacity are seen as the natural product of the interaction between organisms and their ecological niche. As Janus *bifrons* connected inside

with outside, past with future, and peace with war, so perception connects living organisms to their ecological niche.

In the following chapters, we will see that this view of perception can be further extended to explain the logical and metaphysical foundations of natural language and its meaning.

## A Note on Sensorial Experience

Before concluding this chapter, we should observe a potential limitation of the model we have described above and, in particular, of the view of sensory experience it promotes. As we have already stressed, the model provides us with a functional account of sensorial experience—that is, an account of the function sensorial experience performs within the broader architecture of sensorial perception. The function performed by sensorial experience, as we saw, is that of providing the perceiving organism with information about its environmental niche in a way that is useful to the organism’s survival in such environment. This is an explicitly epistemic function that is fully grounded in the external world of natural things as it emerges from the interaction between two natural objects—the perceiving organism and its environment.

There is, indeed, a philosophical and scientific tradition according to which to explain conscious experience it suffices to explain its function. There is, however, also a tradition according to which explaining what conscious experience does is not the same as explaining what it is. Those who adhere to this tradition point out that conscious experience has not only functional but also qualitative properties, often called *qualia*. Even when we have explained in the most refined details the function Kazimir’s conscious experience of the color black performs in the broader architecture of his sensorial relation to his environment, we have still not explained the specific qualitative features of Kazimir’s experience of the color black—that is, to borrow a famous expression from Thomas Nagel, what it “feels like” to see the color black from Kazimir’s subjective perspective.

Explaining what conscious experience is, however, is no easy task. For one thing, conscious experience escapes, by its very nature, any objective scientific treatment, at least according to current scientific standards. On the one hand, science is objective. It aims at providing an account of reality that is free from subjective interference. On the other hand, conscious experience is subjective. It is, in fact, the very core of the notion of subjectivity. But then, how can something that is inherently subjective be given an objective explanation? Borrowing a famous metaphor from William James, trying to study phenomenal consciousness scientifically is somewhat like trying to see into the darkness by turning on the light. According to many, the metaphysical foundations of consciousness remain today one of the most complex unsolved mysteries of our world.

We have no ambition to shed light on this debate. Nonetheless, it is useful to briefly consider how the question of what conscious experience is may affect our

current understanding of sensorial perception. On the one hand, one may argue that the problem of what is the fundamental nature of phenomenal experience is irrelevant to our understanding of the role it plays in perception. As we saw, sensorial experience is a system of arbitrary symbols developed by nature to signal sensory states to the perceiver, much in the same way as words are used by the speakers of a language as arbitrary vehicles of meaning. Typically, what words are made of is irrelevant to the function they perform in the language. In fact, languages can be expressed in different material forms—as spoken sounds, written signs, signed gestures, tactile impressions. The same, then, can be argued for sensorial experience. What is the substance of sensorial experience is an interesting question but irrelevant to the role it plays in the architecture of perceptual systems. On the other hand, one may object that, unless we have a clear understanding of what sort of natural object phenomenal experience is, we cannot truly explain how it can possibly manage to perform its function of bridging the gap between the perceiver and its environment. More precisely, we cannot claim we have provided a truly naturalistic explanation of the relation between a perceiver and its environment unless we have proven that the bridge that connects the two is also made of something that is, ultimately, natural.

All in all, the framework we have presented in this chapter may still be an incomplete one. Yet, if a truly naturalistic account of it is at all possible, a naturalistic understanding of its epistemic function seems a first step in the right direction. We will return to these considerations at the end of Chap. 35, when we will discuss the advantages and limitations of extending the model of perception presented in this chapter to the realm of linguistic meaning.

## References and Remarks

Our current scientific understanding of sensory systems is the result of the efforts of a group of pioneering neurologists and cognitive scientists, originally gravitating around the Massachusetts Institute of Technology and McGill University. These include Jerome Lettvin, Donald Hebb, Warren McCulloch, Walter Pitts, Arthur Rosenblueth, and Norbert Wiener. Starting from the 1940s, these researchers developed the first rigorous computational models of the functioning of individual neurons in the brain and, soon after that, of several complex neural structures. These models were then tested in the following decades in different domains, including that of visual perception, by relying on the new technologies that were becoming available at that time. Pioneers in this field were Horace Barlow, Stephen Kuffler, David Hubel, and Thorsten Wiesel. Hubel and Wiesel's work was especially important in determining the functional architecture of the primary visual cortex. Through their experiments (see in particular Hubel and Wiesel 1962), they were able to determine that the neurons in the primary visual cortex are organized topologically around a specific architecture that is based on the function performed by the different neurons. More precisely, neurons that perform similar computations are clustered



into columns that together collect independent information that is then transmitted for integration to higher regions of the brain. Thanks to the efforts of scholars such as Hubel and Weisel, an overall computational theory of visual perception could be developed in the 1970s, under the guide of Marvin Minsky and Seymour Papert. The most remarkable outcome of this work is, without doubt, David Marr's essay *Vision*, published in 1982, soon after Marr's tragic and sudden death in 1980 (Marr 1982). Our presentation in this chapter was based on Matthen's 2005 essay *Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception* (Matthen 2005). In it, Matthen offers a detailed, comprehensive, and yet accessible overview of the state of the art in the study of the physiology, neurology, and cognition of sensorial perception as well as an original and exciting philosophical interpretation of its implications.

For more information on Barlow's work on feature detection, we refer the reader to Barlow (1972); for evidence on the dissociation and resolution of feature maps, to Masland (2001, 2003).

For experimental evidence on tracking, we refer the reader to Blaser et al. (2000) and Pylyshyn (1989, 2001). The analysis of binding in terms of predication—that is, in terms of attribution of properties to an object—is due to Austen Clark (see Clark 2000).

The notion of “narcissism”, as applied to sensory systems, is due to Kathleen Akins and was first presented in her article “Of Sensory Systems and the ‘Aboutness’ of Mental States” (Akins 1996). In her original article (see especially section IV, “Philosophical Implications”), Akins contends that narcissism ultimately leads to a form of internalism concerning sensory systems as well as a form of anti-representationalism (see especially section IV of her article). Internalism means that sensory system should not be understood in terms of correlations between sensory states and external properties. Anti-representationalism means that sensory system should not be understood as providing a representational content. Akins' conclusions suggest an analysis of sensory systems in the spirit of Fodor's notion of methodological solipsism (Fodor 1980). Methodological solipsism is the thesis that the psychological states of an organism must be understood exclusively on the basis of the states, properties, and relations that are internal to the organism, without any reference to either the physical or societal environment in which the organism is embedded. Methodological solipsism is, therefore, a form of psychological internalism, meant to complement the computational view of the mind, whereby the mind is understood as a system of manipulation of symbols that operates independently of the content of such symbols, and to oppose the representational view of the mind, also often referred to as psychological naturalism, whereby psychological states are understood in terms of their content—what they represent. While we believe the notion of narcissism applies correctly to sensory systems, we disagree with Akin's conclusions and, as we have illustrated in this chapter, endorse Matthen's thesis (Matthen 2005) that sensory perception is both realist—the system is resolutely anchored to the real world of material things and meant to collect information about it—and epistemic—its goal is that of providing information to the organism it serves. On this note, it is interesting to observe how the scientific study of



perception has exerted an important influence in the most recent discussion in the philosophy of mind. Philosophers who have recently developed frameworks aiming at reconciling the idea that perception is representational with a naturalistic view of the mind are Tyler Burge and Susanna Siegel—see, in particular, Burge’s essay *Origins of Objectivity* (Burge 2010) and Siegel’s *The Contents of Visual Experience* (Siegel 2010). Fodor and Pylyshyn’s have recently attempted at providing a purely referentialist—hence externalist—theory of concepts and their semantic content that is based on Pylyshyn’s experiments on tracking. Their theory is presented in *Minds without Meanings: An Essay on the Content of Concepts* (Fodor and Pylyshyn 2015).

The view of sensorial experience as signals of sensory states is due to Matthen himself—see especially chapter 10 of Matthen (2005).

The model of perception we have presented is simplified in a number of ways for the sake of exposition. The simplifications we have introduced, however, do not hinge on the conclusions we can draw from the model or on the use we will make of it in the coming chapters. There are nonetheless two simplifications that are worth briefly mentioning. The first concerns the logic behind the firing of sensory neurons. In the chapter we have presented a *Boolean* approach to this issue, whereby neurons fire if a condition is met in their receptive field and do not fire if the same condition is not met. In reality, the firing of neurons is best described in *Bayesian* terms. Neurons do not deliver the equivalent of a truth value (1/0) depending on whether the condition is or is not met. Rather, they deliver a probability value (a number between 1 and 0) expressing the probability that the relevant condition is met. This means that sensory states are best described within a probability space, rather than a possibility space. On this, see Barlow (1972). The second simplification concerns the process of binding. In this chapter, we have maintained that the binding of features and objects into sensory states is performed automatically and systematically. In fact, binding occurs only under conditions of focal attention. That is, it is only when the perceiver pays attention to a specific location that the features and objects in such location are bound into an integrated sensory state. On this, see Treisman (1996).

Interestingly, the hypothetical condition we have envisaged in this chapter whereby a perceiver experiences red as green and green as red is predicted to exist as the sum of two independent and well attested actual conditions: *protanopia*—a genetic condition that makes perceivers experience the color red as green—and *deuteranopia*—a condition that makes perceiver experience green as red. As the two conditions are independent from one another, there is a concrete probability that they may co-occur. Yet, subjects with such a condition have never been reported. Arguably, this is precisely because such a condition is not detectable on behavioral grounds. On this issue, see Palmer (1999).

Our discussion of color has assumed that color correlates with wave length. This is not entirely correct, as demonstrated by the so-called “Land effect”. Color vision is a visual discrimination capacity that essentially relies on wavelength-discriminating sensors to ground differential responses to light differing in wavelength only. Biologically, we know that there are three different types of cone cells

in the retina, containing different visual pigments that are sensitive to the different portions of the visible light spectrum. The flux of radiant energy from an object is absorbed by the cone cells in accordance with the spectral sensitivities proper to the cone cells. This flux is transduced into a signal that is sent along the optical nerve to the cortex. Now, the Land Effect detected (and long studied, also experimentally) by Edwin H. Land and John J. McCann consists in the apparently puzzling observation that color perception remains constant regardless of light modulation (Land 1959). In other words, the eye—or better to say, the mind—perceives the same colors independently of different illumination conditions. Conventional ideas about color vision are based on the hypothesis that the viewer essentially responds to the radiant energy flux reaching her eye; in this way, color detection should depend on the illumination of a scene. However, the color constancy phenomenon implied by the Land effect seems to require a theory according to which the viewer response is not to the radiant energy flux at any given unit area; the response is rather based on the comparison that the viewer makes between the flux at any given unit area and the flux over the entire field of view, relativized to the three separate wavelength systems. This theory is known as the “retinex theory” (Land 1977). The term “retinex” is a portmanteau from the words “retina” and “cortex”. In a nutshell, the theory contends that three distinct images or lightnesses are compared in order to produce a sensation of color. The comparison is a plotting of three different lightness values, whereby color is determined by the single point at which these values converge. Retinex theory has inspired the creation of computer algorithms intended to imitate the human visual system. From our perspective, the fact, say, that a yellow object is constantly seen as yellow, independently of the conditions at which the object is exposed to light, simply defines a sharp difference between the human visual system and color photography. We think that this difference represents an important argument in favor of Burge’s arguments (Burge 2010) against what he calls “individual representationalism”—that is, the thesis that all of an individual’s representational mental states are essentially independent of the relation to a wider reality (in this respect, Burge’s notion of individual representationalism also encompasses Fodor’s notion of methodological solipsism, mentioned above). Conversely, the anti-individualism promoted by Burge contends that the law-like patterns that define the formation laws of perceptual psychology, leading from given registrations of proximal stimuli to well-defined representational states (as the perceptual states involved in color detection) are the results of phylogenetically determined, that is, evolutionarily justified, processes, though it should also be emphasized that the success of representational states in perception is not immediately reducible to the measure of success in terms of biological function. Crucially, the in-depth study of how the visual system operates on the physical correlates of color sensations makes it perfectly clear that color detection is a function of the mind as a whole. This conclusion is confirmed by the study of higher-order visual effects such as the Checker Shadow optical illusion (see Adelson 2000) and the systematic (though still poorly understood) disagreement among subjects in color sensation (between black and royal blue on one side, and white and gold on the other side), as found in the so-called “dress illusion” (see Lafer-Sousa et al. 2015).

A note of caution is in order concerning on the use of the term “binding”. In the context of the theory of sensory perception we have illustrated in this chapter, the term refers to the process of attributing features to objects. This use of the term must be distinguished from the use that is made of it in linguistic theory where it typically refers to the relation that associates an anaphoric term—such as a pronoun—to its antecedent.

The literature on consciousness is truly vast and encompasses a large plethora of different approaches—behavioral, psychological, neurological, philosophical, biological, physical, mathematical, computational. Amongst the most vocal supporters of the functional approach to consciousness in recent times we should mention Dennett (1991). Amongst those who have recently defended the view that consciousness has qualitative properties that are irreducible to our current understanding of the natural world we should mention David Chalmers (Chalmers 1996, 2010). The expression by Thomas Nagel referenced in the text above is from his article “What is it like to be a bat?” (Nagel 1974).

## References

- Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 339–351). Cambridge, MA: MIT Press.
- Akins, K. (1996). Of sensory systems and the ‘aboutness’ of mental states. *Journal of Philosophy*, 93, 337–372.
- Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception*, 1, 371–394.
- Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature*, 408, 196–199.
- Burge, T. (2010). *Origins of objectivity*. Oxford: Oxford University Press.
- Chalmers, D. (1996). *The conscious mind*. New York/Oxford: Oxford University Press.
- Chalmers, D. (2010). *The character of consciousness*. Oxford: Oxford University Press.
- Clark, A. (2000). *A theory of sentience*. Oxford: Oxford University Press.
- Dennett, C. D. (1991). *Consciousness explained*. Boston: Little Brown.
- Fodor, J. A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences*, 3(1), 63–73.
- Fodor, J. A., & Pylyshyn, Z. (2015). *Minds without meanings: An essay on the content of concepts*. Cambridge, MA: The MIT Press.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of Physiology*, 160, 106–154.
- Lafer-Sousa, R., Hermann, K. L., & Conway, B. R. (2015). Striking individual differences in color perception uncovered by ‘the dress’ photograph. *Current Biology*, 25, 13.
- Land, E. H. (1959). Experiments in color vision. *Scientific American*, 200(5), 84–99.
- Land, E. H. (1977). The retinex theory of color vision. *Scientific American*, 237(6), 108–129.
- Marr, D. (1982). *Vision*. New York: W. H. Freeman.
- Masland, R. H. (2001). Neuronal diversity in the retina. *Current Opinion in Neurobiology*, 11, 431–436.
- Masland, R. H. (2003). Vision: The retina’s fancy tricks. *Nature*, 423, 387–388.
- Matthen, M. (2005). *Seeing, doing, and knowing. A philosophical theory of sense perception*. Oxford: Oxford University Press.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83, 435–450.

- Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT.
- Pylyshyn, Z. (1989). The role of location indexes in spatial perception. *Cognition*, 32, 65–97.
- Pylyshyn, Z. (2001). Visual indexes, preconceptual objecthood, and situated vision. *Cognition*, 80, 127–158.
- Siegel, S. (2010). *The contents of visual experience*. Oxford: Oxford University Press.
- Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, 6, 171–178.

## Chapter 34

# Meaning as Perception (Logic)



### Meaning as Perception

In the final chapters of our book, we shall present and discuss the hypothesis of *meaning as perception*. As we saw, semantic externalism identifies the fundamental properties of natural language meaning in the external world of material things whereas semantic internalism identifies them in the inner realm of the mind. In contrast with both these hypotheses, the thesis of meaning as perception identifies the roots of natural language meaning in perception. Notably, the hypothesis is not that natural language is only and exclusively about the things we perceive. The hypothesis is, rather, that when we talk about things in natural language, we talk about them *as if* they were objects of perception. Indeed, natural language allows us to talk about many things, including things we do not acknowledge through perception. Yet, whatever it is that we talk about, we talk about it as being endowed with the logical and metaphysical properties of objects of perception.

In support of the hypothesis of meaning as perception, we will present two arguments. The first, which we will discuss in this chapter, is that the logic of perception corresponds to a remarkable extent to that of natural language meaning, as it is described in model-theoretic semantics. The second, which we will discuss in the following chapter, is that the hypothesis of meaning as perception offers a valuable framework for explaining the metaphysical properties of natural language meaning and, in particular, for reconciling the roles world and mind play in determining such properties.

## Logic of Meaning and Logic of Perception

It is not surprising to find that, in order to explain the functioning of perception, scientists often resort to linguistic metaphors. In effect, as we already noticed in the previous chapter, there are a number of parallels between the logical framework that emerges from the study of perception and the one that emerges from the study of natural language meaning. As we saw in part I of this book, Montague set the foundations of model-theoretic natural language semantics with the explicit purpose of accounting for the productivity of natural language meaning. Natural language, he observed, allows us to use a finite set of simple meanings in order to produce an infinite set of complex ones. To explain this property of natural language, Montague adopted a strategy already envisaged by Frege whereby simple meanings are combined into complex ones in harmony with the grammatical structure that holds together simple expressions into complex ones. In the remainder of this chapter, we will see that the logical framework at the basis of Montague's theory overlaps in significant ways with the one envisaged by modern scientists to explain perception.

### Objects and Properties

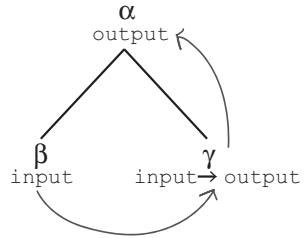
To begin with, both meaning and perception are grounded on two logical primitives, objects and properties. Sensory systems, as we saw in the previous chapter, start off by performing two operations, tracking objects and mapping features – that is, properties of objects. Likewise, the simplest logical elements at the core of Montague's compositional machinery are objects and properties of objects.

### Propositions and Functional Application

Sensory systems proceed then by binding together objects and properties in order to produce sensory states—*de re* propositions, attributing properties to objects. *De re* propositions, attributing properties to objects, are also the simplest form of proposition that can be produced in natural language. Importantly, the process that derives sensory states from objects and properties is logically equivalent to the mechanism of functional application that derives simple *de re* propositions from objects and properties in natural language. As we saw back in Chap. 7, the mechanism of functional application derives the meaning of a complex expression by applying one of its nodes as the input of the function denoted by the other node, as exemplified in Fig. 34.1.

The simplest logical types that satisfy the configuration in Fig. 34.1 are, indeed, objects—the simplest type of argument—and properties—the simplest type of function. As demonstrated by Montague, the notions of object, property, and

**Fig. 34.1** The mechanism of functional application



proposition in combination with the mechanism of functional application are all we need in order to derive the productivity of natural language meaning in all its unbound complexity. These notions mirror with remarkable precision those at the heart of sensory systems, as they are understood by scientists today.

### Reference and Representation

The fact that perception and language both adopt a propositional format in order to package information suggests that they share an even more fundamental logical feature, *referentiality*. Both systems are engaged in the production of *representations*. Their aim is to provide descriptions of a model of reference. In fact, both can be understood as producing structures that are homomorphic to the distal reality they aim to describe.

### Compositionality and Exaptation

Despite these parallels, there is also a significant difference between language and perception. Language does not simply mirror the fundamental logical structure of perception. It further exploits it to a much larger extent. In perception, the combinatorial properties of objects and properties serve the sole purpose of producing sensory states. In language, however, the same mechanism is exploited in order to produce logical structures of much higher—in fact, potentially unbound—complexity. Not only do we use language to attribute properties to objects but also properties of properties to properties, properties of properties of properties to properties of properties, and so on to the degree of complexity we wish.

In this respect, language may offer an interesting case of the evolutionary phenomenon for which the paleontologists Stephen Jay Gould and Elisabeth Vrba coined the term *exaptation*. As is well known, Darwin’s theory of evolution aims at explaining the current features of natural organisms as the product of a process of *adaptation* to their environment. In this context, the term “adaptation” indicates the process whereby a feature of an organism is built through the long process of natural selection for the purpose of performing its current function. It was Darwin himself,

however, who pointed out in his *On the Origin of Species* that there are some problematic exceptions to this paradigm of explanation. As an example, he mentioned the case of the sutures in the skull of young mammals. These sutures are fundamental in assisting baby and mother during parturition by making the baby skull more flexible. One may therefore think that they are the product of adaptation—that is, that they have been shaped by natural selection to perform their current function of aiding parturition. But that, Darwin observes, cannot be the case because we find the same sutures also in the skull of birds and reptiles, which are not born through parturition, but only have to escape an egg. Darwin concluded that we cannot explain this feature as the outcome of adaptation. In fact, he included this example in a chapter aptly entitled “Difficulties on Theory”. In the decades that followed Darwin’s original formulation of his theory, more and more cases emerged of natural features in a variety of organisms that had originally evolved for a different function—or, in some cases, for no particular function at all—and, only later, had been co-opted into performing their current function. A famous example is that of bird feathers, which, contrary to what one may be tempted to think, originally developed to perform a function of thermoregulation and only later were exploited for flying—as demonstrated by the fact that the first prehistoric birds were already covered by a rich mantle of feathers even though they were unable to fly. It took some time before this evolutionary strategy of exploiting existing functions for new tasks would be given recognition within Darwin’s original theoretical framework. It was only in 1982 that Gould and Vrba proposed in their article “Exaptation—A Missing Term in the Science of Form” that we must distinguish two different mechanisms at work in natural evolution, adaptation and exaptation. Adaptation is the process whereby a feature develops through natural selection for its current function. Exaptation is the process whereby a feature that has developed for a particular function—or no function at all—is co-opted into a new function.

The parallel between language and perception suggests a further case of exaptation. The mechanism of composition between objects and properties operates in perception for the sake of a specific and well-delimited function: that of providing sensory states—simple *de re* propositions attributing properties to objects. In language, however, this same mechanism is exploited recursively for the sake of constructing propositions of a significantly higher and, in fact, potentially unbound degree of complexity.

## References and Remarks

The notion of exaptation was first introduced by Gould and Vrba in their article “Exaptation-A Missing Term in the Science of Form” (Gould and Vrba 1982). The exaptation of wings for flight, after their adaptation for thermoregulation, is especially clear in insects. On this, see the classic article by Joel Kingsolver and Mimi Koehl “Aerodynamics, thermoregulation, and the evolution of insect wings: differential scaling and evolutionary change” (Kingsolver and Koehl 1985).



The notion of exaptation has been applied to the domain of language evolution by other scholars. Notable examples are, among others, Robert Berwick and Noam Chomsky (see their *Why only us: Language and evolution*, Berwick and Chomsky 2016), Rudolf Botha (see his article “How much of language, if any, came about in the same sort of way as the brooding chamber in snails?”, Botha 2001), Tecumseh Fitch (see his book *The Evolution of Language*, Fitch 2011), and Roger Lass (see his “How to do things with junk: Exaptation in language evolution”, Lass 1990).

## References

- Botha, R. P. (2001). How much of language, if any, came about in the same sort of way as the brooding chamber in snails? *Language and Communication*, 21(3), 225–243.
- Berwick, R. C., & Chomsky, N. (2016). *Why only us: Language and evolution*. Cambridge, MA: MIT Press.
- Fitch, W. T. (2011). *The evolution of language*. New York, NY: Cambridge University Press.
- Gould, S. J., & Vrba, E. S. (1982). Exaptation—a missing term in the science of form. *Paleobiology*, 8(1), 4–15.
- Kingsolver, J. G., & Koehl, M. A. R. (1985). Aerodynamics, thermoregulation, and the evolution of insect wings: differential scaling and evolutionary change. *Evolution*, 39, 488–504.
- Lass, R. (1990). How to do things with junk: Exaptation in language evolution. *Journal of Linguistics*, 26(1), 79–102.

## Chapter 35

# Meaning as Perception (Metaphysics)



The parallels between the logical structure of perception—as uncovered through the scientific study of sensory systems—and the logical structure of linguistic meaning—as envisaged by model-theoretic semanticists—are remarkable. Notions such as objects, properties, propositions, and compositionality, as they have been identified in natural language through the study of the relationship between meaning and grammar, correspond systematically to those identified by cognitive scientists in the domain of sensorial perception. In this chapter, we shall discuss how the thesis of meaning as perception provides an effective framework for reconstructing the divide between world and mind in natural language meaning.

As we saw, the main lesson to be drawn from the scientific study of sensory systems is that perception is not what divides the mind from the world but, rather, the glue that holds together an organism to its environment. The fact that sensorial information is packaged into propositions, with truth-conditional properties, presented to the conscious mind in the form of conscious experience with phenomenal properties, is not the wicked trick of a malevolent devil. To the contrary, sensory states and sensory experience are the tools nature developed throughout the course of our evolutionary history to provide us with useful information about our environment—information explicitly intended to serve and support us in the challenging task of living and prospering in the specific context of our environmental niche. Like Janus *bifrons* perception has two faces—one oriented outward, towards the world, the other oriented inward, towards the organism. At one end, sensory systems reach out to the world by referring to its objects and classifying its properties. At the other end, they connect to the organism by packaging the information collected in the form of sensory states—*de re* propositions representing objects and their properties—and presenting them in the form of conscious experience. Their function is, ultimately, that of representing the world to inform the observer.

Extending this view of perception to linguistic meaning, as the thesis of meaning as perception suggests, allows us to reconcile the observation that meaning has an undeniable cognitive value with the observation that meaning is inextricably

anchored to the material environment in which it occurs. In the remainder of this chapter, we will consider three domains of linguistic meaning where the hypothesis of meaning as perception finds a useful application.

## Reference as Tracking

In part II, we saw that the way we refer to objects in natural language is shaped by two factors, a material one and an epistemic one. As we saw through Chaps. 21, 22, and 23, linguistic expressions and their reference are bound to their context of use through an indexical function—a relation of material causation linking expressions to the material environment in which they are used. This relation can be direct—as in the case of indexical terms such as “I”, “here”, or “now”—or inherited through the practices of the linguistic community—as in the case of proper names such as “Kazimir Malevich”, “Hesperus”, or “Phosphorus”. As we noted, the observation that meaning has a causal history has been adopted by the advocates of semantic externalism as an argument for the view that meaning is inescapably bound to the material circumstances in which the language is used through a relation of material causation. Indexicality, however, is not the only factor that is relevant to determining the reference of linguistic expressions. As we saw through Chaps. 24, 25, 26, and 27, meaning is not only a function of the material environment in which expressions are used. It is also a function of the way in which speakers acknowledge such environment. Hence, the function that determines meaning is not only causal but also epistemic. This is demonstrated by the fact that natural language allows speakers to refer to one and the same object in different ways, hence contributing different meanings. This is what happens, for example, when speakers refer to themselves in the awareness of doing so—hence contributing *de se* propositions—as opposed to when they do so without awareness of doing so—hence contributing merely *de re* propositions. This is also what allows speakers to attribute contradictory properties to one and the same object, without falling into blunt irrationality—as, for example, when Kazimir acknowledges one and the same figure to be and, at the same time, not to be the depiction of an elephant.

As we saw, the canonical way to account for the epistemic qualities of linguistic meaning capitalizes on the notion of acquaintance. Reference, the story goes, is not solely determined by the material conditions in which language is used but also by the way in which speakers are acquainted with them. The notion of acquaintance, as it is understood in this context, is due to Russell who, in his model of human knowledge, famously distinguished between knowledge by acquaintance—the passive registration of environmental information as it is performed by our senses—and knowledge by description—the conscious mental activity of organizing the information gathered by the senses around properties and relations. According to this view, indexicality and acquaintance are the two pillars upon which linguistic meaning is constructed. Indexicality relates it to the material environment in which language is used. Acquaintance relates it to the perspective of the cognitive agents who

use the language. But how do these two aspects of linguistic meaning relate to one another?

The first thing we should observe is that the notion of acquaintance, as we know it from Russell, is problematic in light of what we know now about perception. The distinction between acquaintance and description is based on what we referred to, in Chap. 32, as the “incorrect” view of perception. Like Descartes, Russell maintains that perception proceeds in two steps. First, the senses passively register the external input. This first stage corresponds to Russell’s notion of acquaintance. Then, the mind conceptualizes it. This second stage corresponds to Russell’s notion of description. We now know that this view is incorrect because perception is not a process of mere sensory registration but, in fact, an act of classification from its very inception.

Upon closer reflection, we can draw a parallel reason for discontent with the notion of acquaintance in the domain of language. In the context of natural language meaning, acquaintance is supposed to perform a function of object identification. It is supposed to allow speakers to individuate objects in their environment so that speakers can refer to them through language. However, as we saw, in the context of the Cartesian view of perception, the notion of object is a derivative one. Objects are not the product of sensorial acquaintance. They are rather singled out from the record provided by the senses through conceptual reflection. This means that the notion of object is a product of the process of description. If this is correct, acquaintance cannot be what allows speakers to individuate objects in their environment and, hence, refer to them through language.

The correct view of perception offers a solution to this problem and, at the same time, provides a unified account of how the epistemic and material sides of linguistic reference relate to one another. The scientific study of perception has taught us that objects are individuated through the mechanism of tracking. Starting from its most peripheral receptors, perception classifies the information from its environment in terms of objects and properties and does so with the explicit purpose of distilling from the environment the information that is useful to the perceiving organism. It is at this early stage that the mechanism of tracking detects the objects in the perceiver’s environment. Tracking allows us to subsume the notions of indexicality and acquaintance within a single mechanism. Objects of tracking are, at once, materially related to the external reality they represent and epistemically related to the cognitive agent they inform. They are meant to capture a feature of the external world with the explicit purpose of informing the organism about their presence.

If we assume, as suggested by the hypothesis of meaning as perception, that the relation that holds together linguistic expressions and their reference is endowed with the same essential properties as tracking, we can naturally account for the peculiar properties of *de re* propositions. As we observed, perception is narcissist. Its function is that of providing useful information—information, that is, that supports the perceiving organism in surviving and prospering in its environmental niche. From this functional perspective, true information is useful information that conforms to the correctness conditions shaped by evolution for the functioning of perceptual systems. In vision, for instance, these conditions typically involve the relation

between the distal reality that the system aims at capturing and the proximal representation that is made of it. Tracking is no exception. In fact, tracking can be unfaithful to objective reality in at least two ways: It can identify a single object as two different ones—when, for example, an object is tracked as different objects under different circumstances; and it can fail to identify two objects as different ones, tracking them as one and the same object. The former type of mistake allows us to explain the peculiar properties of *de re* propositions. If the objects *de re* propositions are about are identified by means of a relation that has the same essential properties as tracking, we can readily explain why a speaker can entertain different—and, sometimes, even contradictory—beliefs about one and the same object. For example, we can explain how a speaker can entertain propositions about herself with and without awareness of doing so. In one case, she is tracking the relevant *res* as herself. In the other case, she is tracking herself as a different *res*. Similarly, we can now explain how a speaker can attribute contradictory properties to the same *res* without being illogical. She is simply tracking one and the same object as two different ones.

Thanks to the notion of tracking we can account for another linguistic puzzle: Putnam's Twin Earth experiment. In Putnam's imaginary scenario, two perfectly identical individuals—actual Kazimir and Twin Kazimir—with perfectly identical brains and, therefore, perfectly identical thoughts, live on two planets—actual Earth and Twin Earth—that are perfectly identical to one another except for the fact that on actual Earth water is H<sub>2</sub>O whereas on Twin Earth water has a different molecular structure. In the two planets, the word “water”, as pronounced by the two Kazimir's, has different meanings because it refers to different substances, which appear the same to the two Kazimir's but, as we know, are altogether different natural objects. The thought experiment demonstrates, as we know, that the meaning of the word “water” has nothing to do with what speakers think of water but, rather, with what water really is. This, as we saw at the beginning of part II, is a cogent theoretical argument for semantic externalism. At end of part II, however, we saw that the internalist view of meaning has also something to say about Putnam's imaginary scenario. In effect, even though the word “water” has different meanings when pronounced by the two Kazimir's, it does mean the same to them. If they are thirsty, they can both ask for a glass of “water” to satisfy their thirst. If they want to wash their clothes, they can both use soap and “water” to do so. Is it possible to reconcile these two aspects of the meaning of “water”? Is it possible, that is, to reconcile the fact that what is true and false of it depends on what water actually is with the fact that the word has the same cognitive value to its speakers despite referring to different objects? Tracking provides a solution to the puzzle. If the reference of “water” is identified through a relation that has the same essential properties of tracking, as we find it in perception, we can explain how the two Kazimir's manage to refer to two altogether different natural objects while, at the same time, attributing to them the same cognitive value. What they are doing is tracking two different objects as one and the same. As we saw, tracking can represent two different objects as one, if the conditions are such that doing so is the most informative way for the sensory system to represent the environment to the perceiving organism. This is what the two Kazimir's are doing with the liquid and colorless substances in their environ-

ment. Even though the two substances are different, the two Kazimir's are tracking them as one and the same. This solution allows us to vindicate, on the one hand, Putnam's observation that what is true or false of the word "water" depends on what water really is and, on the other hand, the internalist's intuition that "water" contributes the same cognitive value to the two Kazimir's.

## Entertaining Propositions and Expressing Propositional Attitudes

Semantic externalism, as we know, puts us in a difficult position when we try to explain how speakers manage to entertain meaning. If meaning belongs to a realm that is strictly separated from the mind, how can it be of any value to the cognitive agents who use it? A most paradigmatic instantiation of this problem concerns the notion of proposition. Frege and Russell held the view that propositions are abstract entities of a platonic sort. They are abstract objects, because immaterial, but that does not make them psychological either, because their existence does not depend on that of a thinking mind. Russell and Frege famously supported this conclusion by observing that propositions are true or false irrespectively of what people think of them. In fact, they are true or false irrespectively of whether there even is someone who thinks of them. That "two plus two makes four" or that "Kazimir Malevich was born in Ukraine" would be true even if nobody thought about it. As we saw in Chap. 19, this view suffers from a radical epistemic problem. If propositions are abstract entities of a platonic sort, how do people manage to grasp them? On the one hand, they cannot be observed like concrete objects, as they are immaterial. On the other hand, they cannot be grasped through thought, as they are not psychological. In the case of mathematical propositions, the problem is typically solved by observing that mathematical propositions are also logical necessities, hence, they are knowable a priori and their truth can be determined on the basis of pure reasoning. This solution, however, does not work for the propositions of natural language, which are, for the most part, contingent. Natural language speakers only rarely use natural language to express absolute necessary truth—doing so would certainly make for some boring conversations. More often than not, natural language speakers find it more informative to talk about facts, rather than abstract truth—the fact that Kazimir Malevich was born in Ukraine, the fact that Pablo Picasso was born in Spain, the fact that Frida is wearing red shoes. These propositions are all true, but not necessarily so—Kazimir Malevich and Pablo Picasso may have been born somewhere else, Frida may have been wearing different shoes. In fact, it is precisely the fact that they are not necessarily true that makes them informative and, therefore, interesting.

In contrast to the abstract view held by Frege and Russell, the study of perception provides a starkly different picture of the notion of proposition. As we saw, perception has an explicit epistemic purpose—that of providing useful information to the perceiving organism about its environment for the sake of its survival. To fulfill this

purpose, sensory systems classify environmental information in terms of objects and properties. Objects and properties are then combined into sensory states, informing the perceiver about which objects have which properties. Sensory states are, hence, equivalent to *de re* propositions attributing properties to objects. In the domain of perception, propositions are not abstract entities, existing in a realm that is independent from the minds of those who acknowledge them. Propositions are, rather, the way information is packaged by sensory systems in order to pass it on to the perceiver, with the specific purpose of informing the perceiver about her natural environment.

Again, perception offers us a useful framework to solve our linguistic puzzle. By extending the view of propositions that emerges from the study of perception to the domain of natural language, we can explain how propositions can be about the external world—the reality that exists independently of the cognitive life of speakers—while, at the same time, retaining informative value for the cognitive agents who entertain them. If the view of meaning as perception is correct, propositions are not abstract objects but the products of a process of classification of the world around objects and properties for the sake of providing useful information to their users—a way of packaging information that, having developed at the interface between speakers and their natural environment, allows them to exchange useful information about a common reality.

There is a further advantage in extending the notion of proposition we find in perception to natural language. Whatever the logical framework that captures the epistemic relation between cognitive agents and the propositions they express in language, that is bound to be a framework that also captures the notion of a propositional attitude. Propositional attitudes, as we saw in Chap. 13, are a fundamental aspect of natural language meaning. They express the relation that holds between cognitive agents—thinkers, believers, hoppers, and dreamers—and propositional contents. As the platonic view of propositions leaves us wondering how propositions can be the objects of the cognitive life of cognitive agents, it also leaves us wondering how language can have the power to express such relation. How does natural language allow us to talk about someone thinking, believing, hoping, or dreaming? Once again, the view of propositions that emerges from the study of perception provides a solution. Propositions, like Janus *bifrons*, stand at the interface between world and cognition, by being about the world while informing the perceiver. Natural language exploits this feature not only to express propositional contents but also to express, recursively, the relation between cognitive agents and the propositional contents they entertain.

## Conscious Experience as Perceptual Experience

Finally, the thesis of meaning as perception provides us with a valuable framework for explaining the specific semantic nature of implicit *de se* reports—those reports where a speaker attributes a psychological state to herself and which are characterized



by immunity to error through misidentification. The specific hypothesis we will consider here is that implicit de se reports are reports of *conscious experience* and conscious experience is strictly intertwined, in the model of interpretation of natural language, with the properties of perception. Our hypothesis is that, as speakers talk about things as if they were objects of perception and properties as if they were perceptual features, so speakers talk about conscious experience as if it was rooted in perceptual experience.

As we saw in Chap. 33, conscious perceptual experience—what perceivers perceive things *as*—is the way devised by nature to signal sensory states to the perceiver for the sake of reasoning and memorization. It is through his perceptual experience, for example, that Kazimir comes to know that there is an object in his visual field that is black and is a square. It is through this experience that he can reason about it and it is this experience what is deposited in his memory.

Perceptual experience, as we observed, has phenomenal properties—properties, that is, that are only arbitrarily related to what the experience represents. The fact that Kazimir sees black the way he does has nothing to do with the physical properties that cause his perception of the color black. Kazimir's experience of the color black is, rather, a symbol decided by nature in the course of the evolution of the human visual sensory system to signal the presence of a certain class of sensory states. Sensorial experience is, henceforth, arbitrary because conventional. Like words in language, perceptual experiences are matched to their corresponding sensory states as a matter of stipulation although, whereas in language the stipulation is made by the community of speakers, in perception it has been made by nature in the course of evolution.

Given these premises, we can explain the peculiar properties of implicit de se reports on the basis of the following two assumptions. The first is the assumption that implicit de se reports are reports of conscious experience. The second is the assumption, suggested by the thesis of meaning as perception, that conscious experience is generally understood as a signal of a natural state of the organism that experiences it. The combination of these two assumptions allows us to reconcile the distinctive subjective nature of de se reports—as highlighted by their immunity to error through misidentification—with the fact that their content is communicable and meaningful to language users other than the speaker.

On the one hand, the distinctive subjective nature of implicit de se reports follows from the fact that they are reports of conscious *phenomenal* experience—what it feels like to subjects from their inner subjective perspective. Sentences such as “I feel cold” and “I am happy” are reports of the speaker's conscious subjective experience of feeling cold and being happy, in the same way as “I see a black square” is a report of the speaker's conscious experience of seeing a black square. In this sense, subjects of perception—the bearers of conscious phenomenal experience—are not the same kind of objects as objects of perception. This is so because, as we have seen in part II, it is impossible to derive the identification of a subject of experience with oneself from any property of natural objects. This view, incidentally, squares with Anscombe's and Recanati's intuition that implicit de se reports do not express de re propositions attributing a property to a *res*.



On the other hand, this view does not entail, as in Anscombe's analysis, that implicit *de se* reports are subjectless nor that, if they have a subject, this subject is a disembodied Cartesian ego—a cognitive agent as detached from the natural world as are the phenomenal properties it experiences. In the framework of sensorial perception we are now acquainted with, conscious phenomenal experience is not detached from the natural world, nor are the agents who experience it. Conscious phenomenal experience is, as we saw, a sign—a conventional symbol established by nature for the explicit purpose of signaling sensory states to the perceiver. Its function is that of providing the perceiving organism with information about its sensory states for the sake of its own survival. Both ends of the epistemic function performed by conscious experience are *natural objects*. On the one side, the information conscious experience provides is a function of the perceiver's natural environment and its own natural state. On the other side, this information is addressed to the perceiver—a natural organism that relies on the information provided to live and thrive in its environmental niche. Within such framework, the occurrence of a phenomenal experience always entails the existence of an organism that experiences it. Granted, an occurrence of "I" in an implicit *de se* report—as in, say, "I am happy"—does not depict a representational state in which a property is ascribed to an object. If this were the case, as it was repeatedly emphasized, there would be no immunity to error through misidentification: The representation might fail to depict things as they are, just because the object we tracked is not the object we think we have tracked. However, no error of misidentification is possible whenever the sentence "I am happy" is uttered truth-faithfully: The sentence is not the result of an act of perception, it is the result of an act of introspection. But does this really show that there is no *res* to which a property is attributed? Not really, if we make the hypothesis that speakers uttering "I am happy" are endowed with (implicit) knowledge that conscious experiences always stand for sensory states as real events in the world, and that sensory states result from the combination of an object (the *res*) with a property. In this way, we can provide the foundations we need to justify Recanati's process of reflection on naturalistic grounds. As we saw in Chap. 29, Recanati expresses the view that implicit *de se* judgments can be turned into explicit *de se* judgements through the inferential process he calls reflection. Through reflection, the awareness of a phenomenal property of a certain subject undergoing an experience can be turned into a property predicated of an object. This happens when the subject of the phenomenal experience comes to recognize the identity between herself—the bearer of the experience—and the object she is in the world. In Chap. 29, we also saw that the process of reflection remains a stipulative solution unless we can justify how it is possible for a disembodied experience to have anything to do with the external world of natural things. The framework we are now considering provides us with such justification as, according to it, experiencing a conscious experience is a sufficient condition to conclude that there is an actual organism—an object of the natural world—experiencing such experience. Reflection is thus not really reflection: it is the unconscious cognitive drift that leads cognitive agents to interpret *pure* experience reports as they were depicting acts of perception. This drift is evolutionarily explained: It originates from the phylogenetically rooted

unconscious knowledge that every experience (crucially including pure experience) is the arbitrary signal of a sensory state—a real event in the real world. It also follows that the inferred *res* is not a disembodied Cartesian ego: It is the object in the world which is identified as the bearer of the experience, as a consequence of modelling introspection—a *pure* experiential state—as a perceptual state of a natural organism.

Within this framework, we can also finally explain how speakers are able to comprehend implicit *de se* reports made by speakers other than themselves as well as how speakers can produce meaningful reports attributing conscious experiences to others. True, a conscious experience can be acknowledged only and exclusively in the privacy of the mind of its experienter. However, in the framework we are considering, the system that anchors conscious experience to natural objects and allows its experienters to acknowledge its significance is common genetic inheritance of all human beings. This means that, although speakers cannot directly acknowledge the conscious experience of others, they can still *infer* it on the basis of their shared nature—the common framework of conventions that anchors conscious phenomenal experience to the behavioral manifestations and the natural properties of the organism that experiences it. Hence, Frida is able to understand Kazimir’s utterance of “I feel cold” because she can infer its meaning on the basis of her own experience of cold and the way this experience commonly manifests itself in the natural world. Similarly, Frida is able to produce the sentence “Kazimir is happy” on the basis of her own experience of happiness and the way her experience regularly correlates to object properties in the natural world.

This is the model of conscious experience and implicit *de se* that we would advocate. Does it solve all the philosophical problems that emerged in the course of our discussion? This is questionable. Undoubtedly, we believe, we should acknowledge a potential limitation of the approach we have proposed. In fact, at the end of Chap. 33, we discussed how the analysis of perceptual experience as a signal may be regarded by some as an incomplete theory of perceptual experience. This is because the theory accounts for the epistemic function performed by perceptual experience but fails to account for its metaphysics—what perceptual experience is made of and whether its essential qualities can be rooted in the natural world. Here, we face a parallel problem. Our framework explains the way we talk about conscious experience on the basis of the epistemic function it performs. It is in reason of the bridge it provides between the natural organism and its natural environment that we are able to treat it as an object of reference in natural language. To this picture, one may object that we have not really provided an account of how conscious experience can provide such a bridge unless we have also explained what conscious experience actually is.

How does the process of reflection allow me to conclude that a phenomenal property—such as that of feeling cold or happy—can be predicated of the individual who accesses the experience? The answer to this question, one may argue, depends on the way we conceive of phenomenal properties in the first place. If we take the position that they cannot be reduced to properties of physical objects—a position that was eloquently defended, among others, by the philosopher David Chalmers—this metaphysical step would be a step too much. According to this view, there is no

serious metaphysical guarantee that phenomenal properties *directly* apply to physical objects. First-person data cannot lead to anything else than verbal reports on subjective experiences. As we know, we can certainly specify—at least in principle—what are the neuronal and computational mechanisms responsible for perceptual discrimination. What we do not properly understand, however, is why a particular neuronal or computational mechanism is associated with the particular sort of perceptual experience it is in fact associated with. The arbitrariness of this process—which we have emphasized above—turns now into *the* metaphysical problem. Granted, seeing red systematically correlates with a specific perceptual mechanism, that is, with a known set of computations involving some specific physical conditions in the external world and some specific neural circuits in the brain. But, as Chalmers correctly emphasizes, if the relevant experience were green instead of red, *nothing* would be actually in need of change in the description of the relevant neural mechanisms and related computations. In other words, any functional description of the mechanisms underlying perception underdetermines the classes of perceptual experience with which they correlate. Or, to put it bluntly, correlation is not explanation. No surprise, since we have explicitly acknowledged that experience is a classificatory device for distinct classes of perceptual states, but the relationship between objective perception and subjective experience is totally arbitrary.

On these grounds, one might conclude that reflection is like a bridge in danger of collapsing. Turning implicit *de se* judgements into explicit *de se* judgements means identifying the subject of a phenomenal experience with the individual—a publicly recognizable object in the world—who has unique access to this experience. However, this identification might be metaphysically unwarranted. After all, it is perhaps more correct to say that phenomenal properties wear their subject on their sleeves, in the sense that the object we are looking for is *implicit* in the unique instantiation of the relevant phenomenal property. As we have seen in the part II, this is actually what implicit *de se* is all about: The object we have knowledge of in an implicit *de se* judgement is immediately and non-reflectively given and does not relate to the phenomenal experience in the way an object relates to the properties that define it.

Of course, nothing prevents the reader from taking issue with Chalmers's view of phenomenal properties. There are in fact eminent philosophers—such as Daniel Dennett—who have long argued that there is nothing to explain in the domain of phenomenal consciousness beyond its function. There are also scholars who have expressed confidence that future developments in neuroscience will be able to close the gap between perception and experience, a gap superbly exemplified, in language, by the contrast between implicit and explicit *de se*, as we have tried to show in this book.

In a sense, we are here at the borders of the known world. Beyond, there is *terra incognita*. Perhaps, we are constitutionally incapable to enter these new mysterious lands. As Noam Chomsky has repeatedly emphasized, from the very moment in which Newton introduced a non-mechanical model of explanation in natural science, there is a sense according to which the body was exorcised, but the ghost remained. In which sense are gravitation laws an explanation of the

relevant phenomena? In which sense is quantum physics an intelligible model of nature? Analogously, in which sense are functional models of the mind intelligible explanations of the way feeling cold, being happy, or perceiving a black square are experienced?

As we saw above, answers might differ, depending on one's philosophical views. From the perspective of this book, we simply intend to show that the thesis of meaning as perception has the potential to solve many conundrums, looking at the cognizing subjects who use language as agents in the natural world. At the same time, we regard it as one of the most important merits of this thesis that it can also elucidate the limits of our present understanding, by pointing to the objective difficulty of getting rid of some residual form of dualism. This should perhaps not come as a surprise, if we pay attention to Chomsky's insight that Cartesian dualism itself arose as part and parcel of a serious core of epistemological preoccupations at the beginning of the Scientific Revolution. The situation has perhaps not radically changed nowadays: In spite of our willingness to give up dualism, doing so clearly entails disregarding a serious core of epistemological preoccupations. It is an important merit of the thesis of meaning as perception to make this consequence fully discernible and appreciable.

## References and Remarks

We refer the readers interested in the issue of phenomenal consciousness to the reference and remarks section of Chap. 33 and the literature cited therein.

## Chapter 36

# Beyond Reason



The thesis of meaning as perception provides us with notions such as object, property, proposition, cognitive agent, conscious experience, and conscious experiencer. In this way, it offers us a valuable framework for capturing the essential ingredients of natural language meaning in a way that reconciles its cognitive role with its material roots. We believe there is a further advantage in adopting the thesis of meaning as perception: It vindicates Wittgenstein's insight that language is an action we perform and that its value resides in the effects it produces. According to the modern view of perception, sensory systems are the tools we, human organisms, use to perform acts of reference to the objects in our natural environment and classification of their properties for the sake of collecting information that is useful to our own survival. The value of such actions is not abstract but practical; not objective but subjective. Their goal is not that of providing perceivers with any objective, abstract truth, but that of being useful to them in the here and now of their existence. If the thesis of meaning as perception is correct, we must conclude as much for language. As an act of reference and classification, its value does not lie in any abstract conception of truth, but in how good it is at providing its users with useful information about their ecology, contributing effectively to the complex of natural relations that binds them with their environment.

Typically, the view of language as action is regarded as incompatible with the following two claims: the claim that language is characterized, at least to some extent, by cognitive properties and the claim that language can be described as a generative system—the unfolding of a set of a priori principles determining its learning, first, and its productive use, then. Within the framework of the theory of meaning as perception, these contradictions no longer subsist.

To begin with, one of the most significant lessons to be drawn from the scientific view of perception is that cognition is also made of actions. Even at the level of their most peripheral receptors, sensory systems perform actions of reference and classification whose products are then employed by the more embedded neural structures to produce propositional contents. In some of his recent works, the philosopher

Scott Soames has advanced the view that propositions are cognitive *operations*, acts performed by cognitive agents either in perception or in cognition for the sake of representing their environment. In their simplest form, these operations are acts of predication, consisting in the attribution of a property to an object. If Soames is correct, then, the idea that cognition is made of actions of predication of properties to objects extends to cognition beyond the domain of perception. In the same spirit of Soames' proposal, the thesis of meaning as perception extends the view of propositions as actions to the domain of natural language and does so without rejecting the view that linguistic actions are performed by cognitive agents for the sake of representing information that is cognitively valuable to them.

The thesis also allows us to resolve the contradiction between the claim that language is an action and the claim that language is the unfolding of a set of a priori principles. Wittgenstein, as we know, relied on his arguments to promote the view that the value of language resides in the here and now of its actual use—the a posteriori effect it elicits from its environment of employment. Within the framework of the thesis of meaning as perception, Wittgenstein's claim is no longer in contradiction with the view that language is grounded on a set of primitive principles. This is possible because, according to the thesis of meaning as perception, the a priori principles that set the essential framework of language learning and use are not of an abstract, metaphysical nature and, in fact, cannot be investigated philosophically on the basis of reasoning alone, nor should their investigation be expected to deliver any metaphysical, transcendental truth. The principles that determine the essential framework of language are, rather, *natural*. They are the product of our evolutionary history. They are the actions we perform in perception and cognition, unconsciously and automatically, as we have been instructed by our genes. There is a substantial difference, here, with the view promoted by Wittgenstein. The value of the linguistic actions we perform does not reside solely and exclusively in the here and now of their performance. It also brings with it the effect it produced in the “there and then” of our ancestors' environment. This is the inheritance that defines the a priori foundations of our ability to learn and use language meaningfully.

## Meaning and Nature

We began our book by observing how Zaum poets and Kazimir Malevich set to revolutionize the canon of the traditional arts by remodeling the very foundations of the notion of creation. The neologism Zaum, as we said, is commonly translated in English as “beyond sense”. Another possible way to translate the morpheme “ym”, however, is “mind”, meaning the realm of reason and rationality. Accordingly, Zaum is sometimes translated as “beyond mind” or “beyond reason” and Zaum poetry is sometimes referred to as “trans-rational poetry”. This particular translation reflects the original aim of Zaum poets to identify the creative foundations of poetry beyond the rational use of language. The same observation applies to Malevich's suprematism. As we saw, Malevich explored what lies beyond the rational use of the

image—the depiction of natural objects—to identify its most essential constitutive ingredients and use them as the foundation of a new, truly creative visual art.

After all, Malevich and his companions may have been right. The way we understand, construct, and talk about the world may very well be the never-ending unfolding of a few irreducible principles. If we are right, however, identifying such principles is no transcendental, trans-rational enterprise. It is neither an exploration into the abstract world of Platonic ideas nor one into the disembodied realm of the Cartesian mind. It is, more likely, an inquiry into nature—ours and that of our environment.

## References and Remarks

Soames' theory of propositions as *operations* of representation performed by cognitive agents either in perception or in cognition is developed in King et al. (2014) and Soames (2007).

It is useful to consider how Pietroski's (2018) position (which we briefly illustrated in the references and remarks section of Chap. 30) compares with the notion of meaning that we have endorsed in our book. The position we have taken is essentially aligned with Soames' claim that propositions cannot be defined as abstract logical structures but should be defined as types of cognitive events of property-ascription. If we omit to do so, we fail to explain how we can correctly represent the world in perception and cognition. On these grounds, we ultimately endorse the view that the set-theoretic notion of proposition cannot be the primitive one and that what ultimately ensures the possibility for sentences in language to be veridical is the notion of proposition that already emerges in perception and, in particular, the veridicality conditions that are associated to individual perceptual events of property-attribution, in strict compliance with the laws of optics and the laws of formation normally assumed in perceptual psychology. However, we also think that it would be wrong to infer from this that propositions cannot be *used* as set-theoretic objects. On the contrary, dealing with propositions as set-theoretic objects establishes the bridge to a quite successful style of cognitive computation—the essentially syntactic computation that characterizes the language of thought and the systems of interpretation in language. In other words, the fact that propositions are rooted in cognitive acts performed by a cognitive agent does not prevent propositions from being endowed with further higher-order properties, which make them computationally tractable within the systems of language and thought. Similarly, it would be preposterous to deny that when propositions are expressed by linguistic expressions, they are endowed with properties *independent* of perception, like hierarchy, recursion and locality. In fact, we fully endorse the familiar position according to which the set-theoretic combinatorics proposed in the composition of non-atomic meanings and the properties of syntax are intimately connected as a foundational property of language.



One of the most prominent empirical questions that arise from the theory of meaning as perception is how different sensory modalities interface with the acquisition of full semantic competence in language. Nowadays, there is a rich array of studies concerning language acquisition in children who are blind or deaf (or blind *and* deaf) from birth, starting with Carol Chomsky's seminal studies on the linguistic abilities of deaf-blind subjects who perceive spoken language through placing a hand on the face of the speaker and monitoring the speaker's articulatory motions, a method of speechreading known as *Tadoma* (see Chomsky 1986). Though the results of these studies illustrate a relatively minor effect of limited language exposure on eventual language achievement, they primarily demonstrate the adequacy of the tactile sense, in these highly trained Tadoma users, for ensuring the development of language and learning to produce speech. Successively, Barbara Gleitman and Lila Landau extensively investigated how experience-deprived individuals can provide evidence for the nature of linguistic—and more specifically semantic—acquisition (see especially Landau and Gleitman 1985; Gleitman, and Landau 2012). These studies confirm that blind (and deaf) children exhibit a full mastery of visual perception predicates such as “look” and “see”, which are in fact among the first words in their vocabulary, as well as linguistic mastery of a variety of color predicates including “red”, “blue” and “orange”. Significantly, it was a common practice of clinicians to counsel parents not to let the children say these words, since they were seen as “empty formalisms - sound without meaning”. Facts are entirely different. It is clear by now that semantic acquisition of “see” and “look” proceeds, for the blind children, from a different set of observational opportunities than is the case with sighted children. More particularly, the relevant information is haptic rather than visual. Blind children being told “to look up” would raise their hands instead of turning their faces upwards. Consistently with these findings, mothers would suggest “to look” at objects that regularly were in the child's hands or at a reasonable touching distance. However, it is not simply the case that blind children offer a physical contact interpretation of the verb “look”. In fact, many more verbs are generally used when the child has a relevant object in hand, such as “give”, “put”, “play” and “hold”. However, the child accurately distinguishes the semantic content of those verbs from the semantic content of perception verbs such as “see” and “look”. In particular, if a blind child is told “you can touch that table but don't look at it”, the child would typically scratch at the table without systematically exploring its surfaces, which she would instead typically do while responding at the command “Now look at it!”. Clearly, then, blind children can easily select the *perceptual content* of “see” and “look”, distinguishing it from non-perceptual predicates such as “hold” and “put”. This is further confirmed by the observation that they are able to conceive color terms as hyponyms of color, and to correctly identify the non-literal interpretation of linguistic expressions such as “green ideas”. Moreover, Landau and Gleitman argue for the existence of syntactic bootstrapping conditions: The fact that perception verbs can select for relative clauses or that-clauses sharply distinguishes them from non-perception verbs such as “hold” and “put”, which do not have these selection properties. All in all, this strongly suggests that semantic acquisition is a resilient cognitive process that essentially abstracts away from the specific



sensory modalities to which certain predicates are superficially tied by effectively selecting a deep notion of perceptual content, contrary to the expectations of British empiricists such as Hume, who explicitly suggested that color words (or even color concepts) could never arise in children who were born blind or deaf. From the perspective we are developing in this book, the finding that the interface between language and perception targets higher-level properties of perception that cut across specific sensory modalities is fascinating in itself, and opens new windows for future research.

## References

- Chomsky, C. (1986). Analytic study of the tadoma method: Language abilities of three deaf-blind subjects. *Journal of Speech and Hearing Research*, 29, 3.
- Gleitman, L. R., & Landau, B. (2012). Every child an isolate: Nature's experiments in language learning. In M. Piattelli-Palmarini & R. Berwick (Eds.), *Rich languages from poor inputs*. Oxford: Oxford University Press.
- King, J. C., Soames, S., & Speaks, J. (2014). *New thinking about propositions*. Oxford: Oxford University Press.
- Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child*. Cambridge, MA: Harvard University Press.
- Pietrosky, P. M. (2018). *Conjoining meanings. Semantics without truth-values*. Oxford: Oxford University Press.
- Soames, S. (2007). *Reference and description. The case against two-dimensionalism*. Princeton: Princeton University Press.

# Author Index

## A

Adamson, S., 178, 183  
Akins, K., 291, 298  
Anscombe, E., 240–246, 249, 250  
Anscombe, G.E.M., 135  
Atkin, A., 173  
Austin, J.L., 140, 145, 174

## B

Bach, E., 53, 117, 122  
Barlow, H.B., 281, 297, 298  
Barwise, J., 75  
Beaver, D., 123  
Bertrand, R., 35, 41  
Berwick, R.C., 307  
Bianchi, V., 113  
Block, N., 264, 266  
Botha, R.P., 307  
Brook, A., 237  
Burge, T., 134, 299, 300

## C

Candlish, S., 161  
Canfield, J.V., 145  
Cantor, G., 42  
Capone, A., 174, 210  
Carnap, R., 95  
Castañeda, H.-N., 210  
Chalmers, D., 192, 301  
Champollion, L., 124  
Chierchia, G., 53, 151, 208, 210  
Chomsky, C., 324

Chomsky, N., 27, 32, 142, 143, 146, 151, 260,  
264, 265, 307  
Church, A., 52  
Clark, A., 298  
Clements, G.N., 208–210  
Condoravdi, C., 123  
Cooper, R., 75  
Cresswell, M.J., 53, 202, 222  
Curry, H.B., 61

## D

Davidson, D., 119, 120, 122, 123  
Delfitto, D., 211  
de Saussure, F., 15  
DeVidi, R.C., 237  
Dowty, D., 52

## E

Espinal, M.T., 23  
Evans, G., 192

## F

Fiengo, R., 192, 193  
Fiorin, G., 211  
Fitch, W.T., 307  
Fodor, J.A., 10, 23, 135, 234, 237, 263, 264,  
266, 298–300  
Frascarelli, V., 113  
Frege, G., 21–23, 74–76, 81, 112,  
127, 154, 155, 157, 173, 175,  
186, 191

**G**

Gallagher, S., 252  
 Gallistel, C.R., 10  
 Garrett, M.F., 10  
 George, W., 161  
 Gombrich, E.H.J., 178, 183  
 Gould, S.J., 305, 306  
 Grice, H.P., 140, 145

**H**

Haas-Spohn, U., 192, 201  
 Harman, G., 250  
 Heim, I., 53, 67, 81  
 Hendriks, H., 76  
 Higginbotham, J., 75, 237, 252, 266  
 Hintikka, J., 95, 106, 107, 109, 110, 112  
 Hubel, D.H., 281, 297

**J**

Jackendoff, R., 111  
 Jackson, F., 192  
 James, W., 252  
 Jespersen, O., 151

**K**

Kamp, H., 81  
 Kanger, S., 95  
 Kaplan, D., 135, 169, 171, 174, 179, 180, 182,  
 183, 187, 188, 190–192, 195–197, 201,  
 210, 211, 222  
 Keenan, E.L., 75  
 King, J.C., 157  
 Kingsolver, J.G., 306  
 Kneale, M., 67  
 Kneale, W.C., 67  
 Knuuttila, S., 102  
 Koehl, M.A.R., 306  
 Kratzer, A., 53, 67, 95  
 Kripke, S.A., 95, 101, 102, 135, 139, 145, 162,  
 187, 188, 190–193, 201

**L**

Land, E.H., 300  
 Larson, R., 75  
 Lass, R., 307  
 Lewis, D., 53, 95, 171–173, 175, 195–197,  
 202, 215, 222  
 Link, G., 151  
 Locke, J., 128

**M**

Maier, E., 202  
 Malcolm, N., 162  
 Marr, D., 298  
 Mateu, J., 23  
 Matthen, M., 277, 282, 291,  
 298, 299  
 May, R., 75, 76, 192, 193  
 McConnell-Ginet, S., 53  
 McGinn, C., 135  
 McGinn, M., 135  
 Mill, J.S., 128  
 Moltmann, F., 122  
 Montague, R., 43, 49, 52, 81  
 Morgan, J.L., 208, 210  
 Mostowski, A., 74

**N**

Nagel, T., 296, 301

**O**

Ogden, C.K., 20

**P**

Parkes, C.H., 10  
 Parsons, T., 53, 123  
 Partee, B., 53, 76  
 Pearson, H., 210, 211  
 Pelletier, F.J., 23, 151  
 Perry, J., 171–173, 175  
 Peters, S., 52  
 Pietroski, P.M., 151, 265  
 Pinker, S., 111  
 Prior, A.N., 95, 122  
 Prosser, S., 237  
 Pryor, J., 235, 237  
 Putnam, H., 129, 132–135, 250  
 Pylyshyn, Z.W., 286, 287, 299

**Q**

Quine, W.V.O., 19, 95, 102, 112,  
 145, 222

**R**

Ramchand, G., 124, 265  
 Reboul, A., 211  
 Recanati, F., 231, 237,  
 246–249

Reichenbach, H., 122  
 Reinhart, T., 193  
 Richards, I.A., 20  
 Rowlands, M., 129, 134  
 Russell, B., 81, 127, 157, 187, 191, 198,  
 199, 201  
 Russell, J., 162  
 Ruys, E.G., 76

**S**

Schlenker, P., 210, 211  
 Schönfinkel, M.I., 62  
 Schützenberger, M.P., 32  
 Searle, J., 140, 145  
 Segal, G., 75  
 Shannon, C.E., 9  
 Shoemaker, S., 236, 251  
 Siegel, S., 299  
 Skinner, B.F., 143, 146  
 Soames, S., 157, 192, 265, 323  
 Speaks, J., 157  
 Stalnaker, R., 95  
 Stalnaker, R.C., 201, 264, 266  
 Stavi, J., 75  
 Stephenson, T.C., 210, 211  
 Strawson, P.F., 81, 140, 145

**T**

Tarski, A., 35, 41, 43

**V**

von Stechow, A., 202, 210, 222  
 von Wright, G.H., 102  
 Vrba, E.S., 305, 306

**W**

Walker, E.C.T., 10  
 Wall, R.E., 52  
 Watson, B., 250  
 Westerstähl, D., 75  
 Whitehead, A.N., 81  
 Wiesel, T.N., 281, 297  
 Winter, J., 123  
 Winter, Y., 76  
 Wittgenstein, L., 19, 20, 35, 41, 127, 129–132,  
 134, 135, 137–140, 142, 143, 145, 146,  
 159–162, 229, 231, 236

**Z**

Zimmerman, T.E., 23  
 Zwarts, J., 123

# General Index

## A

Abstraction, 154, 157  
Abstract logical structures, 323  
Abstract objects, 155  
Acquaintance, 199, 202  
    and context, 214, 218  
    de re and de se beliefs, 221  
    de re reports, 222  
    Kazimir, 214, 221  
    natural language, 239  
    self-acquaintance, 217  
    sensorial apparatus, 239  
    sensorial perception, 230  
Adaptation, 305  
Algebra of entailment, 91–93  
Algebras, 51  
Anti-individualism, 300  
Anti-representationalism, 298  
a posteriori, 99, 102, 190, 191, 255, 322  
a priori, 99, 102, 139, 156, 255, 313, 321, 322  
Attitudes de re, 213, 218, 220–222  
Attitudes de se, *see* De se propositional attitudes  
Autochthonous tradition, 271

## B

Beyond sense  
    artistic expression, 7  
    canon of rules and conventions, 7  
    colors, 9  
    essential ingredients, 8  
    existing language, 7  
    information, 8

    modes of articulation, 8  
    naturalism, 7  
    natural language, 8  
    radical approach, 8  
    visual language, 8  
Binding problem, 284, 285, 301, 304  
*Boolean* approach, 299  
Broad and narrow content, 266  
Broad content, 136

## C

Carol Chomsky's seminal studies, 324  
Cartesian dualism, 319  
Cartesian ego  
    definition, 244  
    disembodied, 317  
    immaterial subject, 245  
    implicit de se judgments, 245  
    independent class, 245  
    introspection, 245  
    motto *cogito ergo sum*, 245  
    ontological subjectivity, 256  
    psychological states, 249  
    semantic externalism, 245  
    Sensory Deprived Kazimir, 244  
Cartesian view of perception, 311  
Centered possible worlds, 215–217  
Centered propositions, 216  
Characteristic function, 47  
Checker Shadow optical illusion, 300  
*Chomsky hierarchy*, 32  
Chomsky's internalist views, 264  
Cognitive events, 265

Cognitive operations, 322  
 Cognitive properties, 321  
 Cognitive skill, 287  
 Color constancy phenomenon, 300  
 Color vision, 299  
 Common noun (CN), 72  
 Community view, 144, 145  
 Complementizer, 106, 107, 112  
 Complex linguistic expression, 21  
 Compositionality, 31, 32, 93, 95  
 Compositional model-theories, 51  
 Comprehensive theory of control, 237  
 Conceptualization, 274  
 Conceptual system, 273  
 Conjoining Meanings, 265  
 Conjunction, 70  
 Conscious experience
 

- actual organism, 316
- black square, 315
- conventions, 317
- de se reports, 315, 317
- epistemic function, 317
- natural objects, 316
- properties, 315
- sensory states, 316

 Content/character, 178, 179, 182, 183, 188–190  
 Contingent, 97, 98  
 Contingently private languages, 130  
 Contradiction, 70  
 Control structure, 233, 234  
 Countable objects, 150  
 Count nouns, 149, 151  
 Cumulativity, 150, 151  
 Curryng, 63–65, 67, 72

**D**

Darwin's theory of evolution, 261  
 Davidsonian event-semantics, 266  
 Declarative sentences, 45, 46  
 De dicto propositional attitude, 110, 111  
 De dicto propositions, 99–102  
 Definite description, 79, 80, 166, 167, 170, 171, 187  
 Definite determiner, 169  
 Demonstratives, 145, 174, 192, 201, 210  
 De re beliefs, 218, 220, 221  
 De re proposition, 99–102, 110, 112  
 Descartes' philosophical preoccupations, 276  
 Descartes' skeptical argument, 241, 242  
 Descartes' skepticism, 290  
 Descriptive approach, 80, 167, 169

De se beliefs, 217, 221  
 De se propositional attitudes
 

- acquaintance, 221
- analysis, 218
- centered possible world, 222
- context and acquaintance, 214
- vs. de re, 195, 196, 205, 206
- first-person pronoun, 195, 204, 205
- free indirect speech, 211
- grammatical control, 207, 208
- Hintikka's semantics, 217
- indirect discourse, 210
- logophoric pronouns, 208–210, 214
- natural language, 203, 213
- truth-conditions, 218, 222
- reference, 204
- self-acquaintance, 205
- value and logic of, 206

 De se sentences, 231  
 Determiner phrase (DP), 30  
 Determiners, 69, 70, 79  
*Deuteranopia*, 299  
 "Difficulties on Theory", 306  
 Divisiveness, 150, 151  
 Dress illusion, 300

**E**

Empty formalisms, 324  
 Environmental niche, 261, 276  
 Epistemic possibilities, 266  
 Epistemic subjectivity, 161, 226  
 Epistemology, 99  
 Error through misidentification, 229, 230  
 "Event-modification problem", 123  
 Events, 119–121  
 Evidentiality, 235, 236  
 Evidential "noda" obligatory, 236  
 Exaptation, 305, 306  
 Explicit model-theory, 41

**F**

Feature mapping, 283, 284, 288  
 Finite-state grammar, 26, 27  
 First-person data, 318  
 First-person pronoun, 163, 171, 195, 204
 

- immunity to error through misidentification, 240
- linguistic perspective, 248
- non-referring, 250, 251
- semantic externalism, 249
- systematic ambiguity, 247

Formal analysis, 127  
 Formal language, 35–41, 43, 59, 60, 83  
 Formal languages interpretation, 256  
 Formal object, 32  
 Frame rate, 286  
 Free indirect speech, 211  
 Frege's distinction, 262, 263  
 Frege's puzzle, 187–190, 192, 193  
 Functional application, 48–52, 72  
 Functions of functions, 62–65

## G

Generalized quantifiers, 74  
 Geometric form, 127  
 Grammar, 25, 38  
 Grammatical control, 207, 208  
 Grammatical relations, 15–17  
 Grammatical sentence, 26  
 Grammatical structures, 21, 207  
 Grounding meaning, 258

## H

Hand-crafted animations, 286  
 “Hesperus”, 186–191, 193  
 Higginbotham's analysis, 237  
 Higginbotham's theory, 252  
 Hintikka's semantics  
   analysis, 214  
   centered propositions, 216, 217  
   de re propositional attitudes, 213  
   de se beliefs, 221  
   de se propositional attitudes, 213, 217, 218  
   Kaplan's distinction, 214  
   Kazimir's first-personal acquaintance, 214  
   reformulations, 217  
 Hintikka's theory, 213  
 Homomorphism, 43, 115, 258  
 Human motor system, 261  
 Hypothesis, 2

## I

Identical environments, 258  
 Image view of perception, 272  
   ambiguity, 274  
   classificatory notions, 273  
   Descartes' idea, 273  
   face-vase illusion, 274  
   indexical relation, 287  
   objects, 287  
   problems, 276, 277

  sensory systems, 273, 274  
   steps, 285  
   stereopsis, 273  
   two-dimensional image, 273  
 Immunity to error through misidentification  
   Anscombe's proposal, 243  
   complementary factors, 243  
   conditions, 232  
   de se propositions, 231  
   de se statements, 231  
   first-person pronoun, 244  
   gerundive and infinitival complements, 237  
   introspection, 232, 239, 241, 251  
   “Kazimir feels cold”, 232  
   “self-reference and self-awareness”, 236  
   semantic externalism, 240  
 Implicit de se  
   Cartesian ego, 245  
   error through misidentification, 233  
   event semantics framework, 237  
   evidentiality, 235  
   existential entailment, 248  
   Higginbotham's theory, 252  
   immunity to error, 231, 242  
   judgments, 240, 242, 246, 249  
   “Perspectival Thought (book)”, 237  
   propositional attitudes, 233–235  
   sensorial acquaintance, 229  
   sentences and reports, 239  
 Indexical expression, 288  
 Indexicality, 310  
 Indexical relations, 287  
 Indexicals  
   a priori statements, 170  
   character, 182, 183  
   content, 182, 183  
   demonstratives, 174  
   as descriptions, 166–168, 170, 171, 173  
   epistemic value, 164  
   expressions, 163, 173  
   first-person, 163, 171, 173  
   function of language, 203  
   Kaplan's logic, 185  
   logical necessity, 172  
   material environment, 185  
   meaning, 183, 185  
   natural languages, 163  
   performatives, 174  
   reference, 164, 166, 173  
   reflexivity, 169  
   self-acquaintance, 173  
   semantic externalism vs. semantic  
     internalism, 164

Individual representationalism, 300  
 “Individuative” quantifiers, 149  
 “I”, non-referring expression, 242–244  
 Interpretation machine, 39, 47  
 Intersection, 70  
 Intransitive verbs, 44, 45, 52  
 Introspection, 227, 232

## J

Janus *bifrons*, 272, 314  
 Japanese grammatical structure, 235

## K

Kaplan logic of demonstratives, 180–182  
 Kazimir, 220, 221  
 “Kazimir Malevich”, 163, 164, 240,  
 310, 313  
 Kazimir’s visual sensory system, 291  
 Knowledge, 99  
 Knowledge by acquaintance, 197–199, 201  
 Knowledge by description, 198, 199, 201  
 Kuleshov effect, 17, 18

## L

Land effect, 299  
 Language acquisition, 146  
 Language as a homomorphism, 40, 41  
 Language as action, 321  
 Language as a mirror, 40, 41  
 Language as use, 143  
 Language learning, 142, 143  
 Language of models, 37–39  
 “*The Language of Thought* (Book)”, 237  
 Language–private, 255  
 Leibniz’s Law, 111  
 Lexicon, 37  
 Linguistic expressions, 14, 83  
 Linguistic meaning, 1, 153, 264  
 Logical framework, 52, 314  
 Logical identity, 186–188  
 Logically private language, 130, 137, 144  
 Logical relations, 1  
 Logophoric pronouns, 208–210

## M

Malevich’s suprematism, 322  
 Mass nouns, 149, 150  
 Material object, 147  
 Mathematical dimension, 177, 178  
 Mathematical objects, 155–157

Mathematical quantities, 37, 116, 153  
 Meaning as an emergent property, 15, 16  
*The Meaning of Meaning*, 20, 135, 262  
 Meaning as nothing, 15, 16, 141  
 Meaning as perception  
   compositionality and exaptation,  
     305, 306  
   conscious experience, 314–319  
   hypothesis, 303  
   Janus *bifrons* perception, 309  
   linguistic meaning, 309  
   logics, 304  
   natural language, 309  
   notions, 321  
   objects and properties, 304  
   propositional attitudes, 313, 314  
   propositions and functional  
     application, 304  
   reference, tracking, 310–313  
   referentiality, 305  
   representations, 305  
   semantic internalism, 303  
   truth-conditional properties, 309  
 Meaning as reference, 13–15, 17–19  
 Meaning as something, 13, 14  
 Meaning as use, 139–141  
 Members, 46  
 Mental objects, 128  
 Merge operation, 29  
 Metaphysics, 117  
*Mind and World* (Book), 276  
 Modal base, 109, 110, 112  
 Modality, 101  
 Model of interpretation, 35, 36, 40  
 Mode of grammatical combination, 25  
 Models, 36, 37  
 Model-theoretic natural language  
   semantics, 51  
 Model-theory, 57, 58, 70, 71  
 Montague grammar, 1, 51, 52, 72–74, 115,  
 116, 121  
 Montague’s approach, 83, 121, 122  
 Montague’s compositional machinery, 304  
 Montague’s experiment, 43  
 Montague’s semantics, 66, 67

## N

Narcissism, 298  
 Narrow contents, 136  
 Natural language, 8, 9, 75, 111, 128, 147, 153,  
 156, 159, 161, 201, 272, 304  
 Natural language expressions, 83  
 Natural language meaning, 7–10, 18, 22, 23



- Natural language metaphysics  
 adverbial modification, 123  
 contingent properties, 117  
 essential properties, 117, 118  
 events, 119–123  
 vs. metaphysics proper, 121, 122  
 necessary properties, 117  
 Occam's razor, 118  
 Prior's logic, 122  
 quality, 115, 116  
 quantitative approach, 118  
 quantity, 115, 116  
 time, 118, 119
- Natural languages, 1, 2, 35, 60, 61, 163
- Natural language sentences, 17
- Natural language syntax, 25, 27, 31
- Necessary, 97, 98
- Neologism, 322
- "Noda", 235
- Non-mechanical model of explanation, 318
- Non-perceptual predicates, 324
- Non-trivial mathematical theory, 42
- Notion of acquaintance, 225
- Noun phrase (NP), 72, 73
- Numeral quantifiers, 149
- O**
- Objectivity, 154
- Object tracking mechanism, 288
- Objects of reference, 14
- Ontological subjectivity, 161, 226, 227
- Ontology, 161
- P**
- "Paderewski's puzzle", 193
- Paradigmatic relations, 15, 16
- Parallel anti-extensionalist, 265
- Passive activity, 226
- Perception, 2
- Perceptual experience, 315, 318
- Perceptual mechanism, 318
- Performatives, 174
- Perspective, 177  
 context, 178  
 mathematical principles, 177, 178  
 psychological principles, 177, 178
- "Phosphorus", 186, 187, 189, 190, 193
- Phrase-structure grammar, 27–31, 105, 106
- Physiology and psychology of perception, 226
- Platonic idealism  
 abstraction, 154, 157  
 abstract objects, 155  
 a priori knowledge, 156  
 concrete objects, 155  
 formal language, 154, 155  
 linguistic meaning, 155, 156  
 mathematical objects, 155–157  
 objectivity, 154  
 semantic externalism, 154
- Points in time, 119
- Possibility space, 108, 109
- Possible states, 87
- Possible worlds, 93, 95, 215
- Possible world semantics, 86–89,  
 109, 116
- Pragmatics, 145
- Principle of compositionality, 18,  
 21–23, 25, 61
- Private language argument, 130–132, 135,  
 137, 138, 159, 160
- Private objects, 160
- Probability, 86
- Proper metaphysics, 121
- Proper names, 13, 43, 44, 69  
 content/character, 188–190, 192  
 contingent a priori, 190, 191  
 vs. definite descriptions, 166  
 as descriptions, 187, 188  
 descriptive theory, 191  
 "Hesperus", 186  
 Kaplan's logic, 186  
 meaning, 192  
 necessary a posteriori, 187, 190, 191  
 "Phosphorus", 186  
 reference, 192
- Properties, 46–48
- Propositional attitudes, 314
- Propositions, 89–91, 97, 105, 108, 156
- Propositions as actions, 322
- Protanopia*, 299
- Psychological dimensions, 177
- Psychological states, 232, 243, 247, 263
- Public languages, 138, 144
- Putnam's imaginary scenario, 312
- Putnam's Twin-Earth experiment, 258
- Puzzling sentences, 169
- Polyshyn's experiments, 287, 299
- Q**
- Quantification  
 determiners, 69  
 model-theory, 70, 71  
 Montague grammar, 72–74
- Quantifiers, 81
- Quine's double-vision problem, 218, 221, 222

**R**

Radical translation, 19  
 Reality, 155  
 Reference, 164, 166, 191  
 Reference, tracking  
   acquaintance, 310, 311  
   awareness, 310  
   contradictory properties, 312  
   factors, 310  
   indexicality, 311  
   Kazimir Malevich, 310  
   linguistic, 311  
   object identification, 311  
   objective reality, 312  
   Putnam's imaginary scenario, 312  
   semantic externalism, 312  
   sensory system, 312  
 Referential expression, 244  
 Referentiality, 305  
 Referring expression, 244  
 Reflection, 316  
 Reflexivity, 169  
 Relations  
   formal language, 59, 60  
   model-theory, 57, 58  
   Montague's semantics, 66, 67  
   natural languages, 60, 61  
 Retinex theory, 300  
 Rubin's illusion, 275  
 Rules of use, 131  
 Russell's definition of acquaintance, 226

**S**

*Salva veritate*, 112  
 Second-person pronoun, 165  
 Self-acquaintance, 173, 178, 186, 199–202,  
   217, 218  
 Self-identification, 197  
 Self-reflective judgements, 232  
 Semantic acquisition, 324  
 Semantic externalism, 2, 128, 129, 132, 135,  
   147, 148, 151, 153, 154, 225, 233,  
   236, 237, 251  
   cognitive relation, 256  
   internalism, 257  
   linguistic meaning, 257  
   natural language expressions, 257  
   ontological subjectivity, 256  
   private languages, 255  
   way speakers, 256  
   weak, 257  
 Semantic internalism, 2, 271

  advantage, 257  
   formidable argument, 259  
   human skill, 260  
   Kazimir, 259  
   linguistic meaning, 260, 262  
   mental activity, 257  
   mental object, 257  
   physiological and cognitive, 260  
   problems, 258  
   Putnam's Twin-Earth experiment, 258  
   structuralist theory, 258  
   "water", 259  
   weak, 260, 261  
 Sensations, 130, 162  
 Sensorial acquaintance, 226, 229, 232  
 Sensorial experience  
   arbitrary relation, 293  
   arbitrary symbols, 297  
   congenital condition, 292  
   convention, 293  
   features  
     active, 295  
     compositional, 295  
     epimistic, 295  
     functional, 294  
     logical, 295  
     proportional, 295  
     realist, 295  
     vivid, 295  
   interpretation, 292  
   Janus *bifrons*, 295  
   Kazimir, 291  
   limitation, 296  
   perception, 291  
   perceptual systems architecture, 297  
   phenomenal object, 292  
   philosophical and scientific tradition, 296  
   sensory states, 293, 294  
   skepticism, 292  
 Sensory Deprived Kazimir, 242, 244  
 Sensory perception, 272  
 Sensory states, 304, 314  
   compositional nature, 289  
   de re propositions, 289, 291, 294  
   objects, 288  
   probability space, 299  
   process, 289  
   properties, 290  
   representational and truth-conditional, 289  
 Sensory systems, 2, 309  
   anti-representationalism, 298  
   binding problem, 284, 285  
   explicit function, 282

feature-detection, 283  
 feature map, 283, 284  
 filters, 279  
 narcissism, 298  
 perceiving objects, 285–288  
 perceiving propositions, 288–290  
 sensorial experience, 291–294  
 types, 279  
 veridicality vs. narcissism, 290, 291  
 visual, 279  
 Sentences, 167  
 Set of pairs, 58  
 Sets, 46–48  
 Sets of objects, 58  
 Set-theoretic combinatorics, 323  
 Set-theoretic objects, 323  
 Set-theoretic relation, 70  
 Set-theory, 42, 43  
 Simple expressions, 25  
 Simple linguistic expressions, 21  
*Sinn/Bedeutung*, 262  
 Soames' proposal, 322  
 Soames' theory, 323  
 Speech acts, 140  
 Standard Theory, 32  
 Stereopsis, 273  
 Structuralism, 17  
 Structuralist theory, 258  
 Subjectivity in language  
   epistemic, 161  
   natural language, 160, 161  
   ontological, 161  
   type of knowledge, 160  
 Super-models, 179, 180, 214, 216  
 Suprematism, 13  
 Symbolization, 262  
 Syntactic bootstrapping conditions, 324  
 Syntagmatic relations, 15, 16  
 Syntax, 25

**T**

*Tadoma*, 324  
 Temperature, 15

*Terra incognita*, 318  
 Theaetetus, 19  
 “Three models for the description of  
   language”, 32  
 Theory of language, 160  
 Theory of meaning, 141, 142  
 Thermometer, 226  
 Thermoregulation, 306  
 Thought experiment, 132  
 Time, 118, 119, 121  
 Tracking, 286, 287  
 Transduction, 274  
 Transformation, 33  
 Transitive verbs, 57, 60, 72  
 Trans-rational poetry, 322  
 Truth-conditions, 85, 86, 214, 217,  
   221, 222  
 Truth-values, 39, 70, 75, 83–85  
 Two-dimensional semantics, 183  
 Two-folded perspective, 225

**V**

Verb phrase (VP), 61, 105, 108  
 Verbs of propositional attitude  
   “believe”, 106, 107  
   context, 112  
   linguistic theory, 111  
   modal base, 112  
   relations, 106, 107  
   semantics, 112  
 Visual arts, 225  
 Visual perception predicates, 324  
 Visual sensory system  
   flat reproduction, 286  
   goal, 280  
   image view of perception, 281  
   neurons, 281  
   traditional view of perception, 282

**W**

Water, 259  
 Well-defined sets, 47