Universiteit
Leiden
The Netherlands

The lazy mindreader : a humanities perspective on mindreading and multiple-order intentionality

Duijn, M.J. van

Citation
Duijn, M. J. van. (2016, April 20). *The lazy mindreader : a humanities perspective on mindreading and multiple-order intentionality*. Retrieved from https://hdl.handle.net/1887/38817

Version:	Corrected Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/38817

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/38817 holds various files of this Leiden University dissertation

Author: Duijn, Max van
Title: The lazy mindreader : a humanities perspective on mindreading and multiple-order intentionality
Issue Date: 2016-04-20

The Lazy Mindreader

A Humanities Perspective on Mindreading and Multiple-Order Intentionality

The PhD research reported in this thesis was funded by the NWO Spinoza prize awarded to Ineke Sluiter in 2010.

Print: CPI / Koninklijke Wöhrman

Copyright © 2016, Max Johannes van Duijn

The Lazy Mindreader

A Humanities Perspective on Mindreading and Multiple-Order Intentionality

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker, volgens besluit van
het College voor Promoties
te verdedigen op woensdag 20 april 2016
klokke 15.00 uur

door

Max Johannes van Duijn
geboren te Leiden
in 1984

Promotores:
Prof.dr. Ineke Sluiter
Prof.dr. Arie Verhagen

Promotiecommissie:
Prof.dr. Herbert H. Clark (Stanford University)
Prof.dr. Robin I.M. Dunbar (University of Oxford)
Prof.dr. Anthonya Visser

Contents

Introduction 9
Reading Guide 20
1. The bigger picture: language, narrative, and social cognition 23
1.1 Mindreading and intentionality 24
1.1.1 Mindreading 24
1.1.2 Intentional states 29
1.1.3 The intentional stance 35
1.1.4 The inferential process: theory, simulation, narrative practice 37
1.2 Embedded mindstates 48
1.2.1 Multiple-order intentionality 48
1.2.2 The roles of language 51
1.3 The social brain 58
1.3.1 Early primate roots 58
1.3.2 The "bonding gap" 60
1.3.3 Cognition and primate social life 63
1.3.4 Mindreading, coordination, and group size 65
1.3.5 Mindreading and social learning 68
1.3.6 Mindreading, language, and narrative 69
2. When narrative takes over: embedded perspectives in Shakespeare's *Othello* 75
2.1. Introduction 76
2.2. Mindreading, Theory of Mind, and Multiple-Order Intentionality 77
2.3. Representing embedded mindstates: from sentence to narrative 80
2.4. The Expository Strategies of Narrative 83
2.4.1 Characterisation 84
2.4.2 Focalisation and viewpoint alternation 87
2.4.3 Framing 89
2.4.4 Episodic structuring 90
2.4.5 Time management 90
2.4.6 Redundancy 91
2.5. Conclusions and Discussion 95
2.5.1 Viewpoint all the way down 95
2.5.2 Context and directions 96

3. Not afraid of Virginia Woolf: embedding and polyphony in the novel 99
3.1 Introduction 100
3.2 Mindstates in literature 102
3.2.1 An early cognitive literary scholar 102
3.2.2 Zunshine's "sociocognitive complexity" 105
3.2.3 The broader picture: coordination of embedded viewpoints 107
3.2.4 A first example: embedding of perspectives in a novel 109
3.3 Viewpoint layers in Mrs Dalloway 111
3.3.1 The first three layers 113
3.3.2 Shifting perspectives 114
3.3.3 Narrative spaces and blending 117
3.3.4 Back to Mrs Dalloway 121
3.4 Discussion and concluding remarks 125
4. Viewpoint packages: tools for communicating and processing complex thoughtscapes 129
4.1 Introduction 130
4.2 Viewpoint, mindstates, and the human ability to read minds 133
4.3 Constructing complexity 134
4.3.1 Compositional complexity: a literary example 134
4.3.2 Compositional complexity: an example from the news 138
4.3.3 Holistic complexity: viewpoint packages 139
4.3.4 Alleged(ly): hedging or coordination of viewpoints? 141
4.3.5 Representing the thoughtscape 143
4.3.6. Viewpoint packages, the emerging story, and (de)compression 145
4.4 Conclusion and discussion 150
4.4.1 Future directions: processing and acquisition 150
4.4.2 The refined definition of viewpoint packages 153
5. Language and joint intentionality: why reflecting on orders of intentionality is the exception, not the default when communicating 155
5.1 Introduction 155
5.2 Association, ostension, and shared intentionality 158
5.2.1 Scott-Phillips' two models of communication 159
5.2.2 Cognitive requirements of ostensive-inferential communication 162
5.2.3 Individual versus shared intentionality 166
5.3 Coordinating mindstates in discourse 177
5.3.1 Dyadic and triadic communication 178
5.3.2 Speaker commitment and viewpoint embedding 181
5.3.3. The general model 187

5.3.3 Thoughtscapes and the model 190
5.3.4 Updating the common ground 196
5.4 Discussion and concluding remarks 198
6. Multiple-order intentionality revisited: analysis of the "mentalising paradigm" 201
6.1 Five central conclusions of the mentalising-test paradigm 202
6.2 Testing mentalising competence 207
6.2.1 Narratives and propositions 211
6.2.2 Packages and inferences drawn from the common ground 219
6.2.3 Judging facts and intentional states 224
6.2.4 True and false statements 230
6.2.5 Structure of interaction 233
6.3 Concluding remarks 236
Conclusion 241
References 257
Samenvatting 271
Acknowledgements 289
Curriculum Vitae 293

Introduction

Imagine: at the pinnacle of the Cold War, when distrust between East and West has congealed into a paralysing paranoia, an engineer coincidentally acquires some knowledge about USA plans to develop a secret weapon. He is persuaded into selling this knowledge to Russian authorities, thus becoming a Soviet spy. However, after a while, USA officials find out about his position and urge him to report periodically what exactly the Soviets know about the plans for the weapon, thus turning him into a double agent. Suddenly the engineer sees a chance to gain a fortune: he runs back to the Soviets and sells them the information *that the Americans know about them knowing about the weapon*. Subsequently, he makes a good impression with the American authorities by telling them that the Soviets now know that they (the Americans) know that the Soviets know about the weapon. This information, that the Americans know that the Soviets know that the Americans know that the Soviets know about the weapon, can then again be sold to the Soviets, and so on…

Cargile, who cites a version of this jest in a philosophical note in *Analysis*, remarks drily: "it seems that he is set up for life, and that the bureaucrats are pretty silly" (1970: 151; the idea was originally based on a comedy routine by Peter Ustinov). While the world of Cold War espionage indeed seems to provide a natural backdrop for the satirical exploration of such complexes of embedded knowledge states, a very similar scenario can also be implemented in a daily-life setting. This is demonstrated by an episode from the fifth season of the popular television sitcom *Friends*. In the episode, the characters Phoebe and Rachel play a practical joke on their friends Monica and Chandler after learning that the two are secretly dating. Phoebe and Rachel first try to keep the discovery to themselves, but through the not-so-clever character of Joey the fresh couple hear that the others know about them dating before the joke has crystallised into something concrete. Monica and Chandler, in turn, start devising a "counter-plot" based on the premise that Phoebe and Rachel *do not know* that they (the couple) know that Phoebe and Rachel know about the dating and are plotting against them. However, again through a clumsy

intervention by Joey, Phoebe and Rachel find out about the preparations for the counter-plot, which triggers them to come up with a cunning counter-counter-plot. Phoebe speaks a line that has long dominated the charts of all-time most renowned sitcom quotes: "*but they don't know we know they know we know*". Rachel immediately grasps what she means, but Joey's face looks dazzled, and when he is asked not to tell anything to the others, the scene ends with him saying: "even if I wanted to...".[1]

What the espionage parody and the episode from *Friends* have in common is that they stage several iterations of knowledge states about knowledge states, variously referred to as *embedded intentional states* or *multiple-order intentionality*: "A knows that B knows that A knows that B...". In this way, a sequence is built up that exhibits a particular type of *recursion*: each constituent added to the left embeds all the constituents to its right. Logically speaking, an infinite regress is lurking, but in practice such a sequence quickly becomes opaque and begins to sound hilarious after just a few steps. This is what the satirist and the *Friends* script writer were relying on for their hilarious effects: while both scenarios are initially realistic, the mind-blowing possibility of infinity promptly comes into view by taking things just a bit too far.

It has been argued extensively that in all sorts of situations of everyday social and cultural life we have to deal with embedded intentional states: we factor into (the planning of) our own future behaviour what we think that others think, know, believe, intend, desire, etc., *including* what we think that they think that we think, and what we think that they think that others think. Planning a dinner party, or gossiping about a neighbour's daughter's new acquaintance, to mention just two examples, arguably involves reasoning about all kinds of intentional states embedded into one another. And even without the particular complexities posed by such activities as event planning or gossipy conversations, we constantly have to deal with embedded intentional states––at least, that is what a wide array of philosophers and researchers working on this topic argue. For example, cognitive neuroscientists have claimed that in order to complete even a basic non-linguistic cooperative task one has to *believe* that the other *believes* that one *intends* to cooperate. Linguists and philosophers of

[1] See episode 14 of season 5, "The One Where Everybody Finds Out" (1999), at around 14.30 minutes.

language have suggested that even basic linguistic communication requires interlocutors to surpass this level, since one has to *understand* that the other *intends* one to *know* that the other *intends* one to *believe* that something is the case. Such arguments have been made for further aspects of everyday socio-cultural living, such as exhibiting moral reasoning and taking part in a religious community (I *want* you to *understand* that God *knows* that we *intend*...), or for appreciating and producing stories (the author *wants* me to *understand* that character A *believes* that character B *hopes* that character C *will believe*...).[2]

The examples given so far, including the two humoristic ones, seem to suggest a paradox: on the one hand, dealing with embedded intentional states has been argued to be indispensible for even basic interactions within our social and cultural environments, whereas on the other hand, a scenario of the form "A knows that B knows that A knows that B knows..." looks so puzzling after only a few steps that it even attracts writers of comedy and satire for its hilarious opacity. Pursuing this line of thought, a series of observations and questions can be put forward:

(1) The complexity involved in dealing with intentional states of others and ourselves, as arguably required by all kinds of situations from daily social and cultural life, has generally been conceptualised as a series of *embedded layers*. Where did this conceptualisation come from? How was it justified? What are its alternatives? And can such alternatives provide other measures of complexity than the number of embedded layers?

(2) Within research traditions of philosophy, psychology, ethology, and the cognitive sciences more broadly, the focus has often been on the

[2] All of this will be introduced in more detail in Chapter 1. For cooperation and multiple-order intentionality see, for instance, Yoshida et al. (2011). The argument that linguistic communication requires dealing with multiple orders of intentionality is rooted in Grice (1957; 1969) and was further developed in Bennett (1976), Sperber (1994; 2000), Papp (2006), and Scott-Phillips (2015). For multiple-order intentionality in relation to religion see Dunbar (2003; 2008) and Dunbar, Gamble, and Gowlett (2010), in relation to morality see Shultz & Dunbar (2007), and in relation to literature see Dunbar (2005), Zunshine (2006), Corballis (2011), and Carney et al. (2014). See also chapter 6 of Dennett's *The intentional stance* (1987), which is a reworked version of his foundational essay 'Intentional systems', describing the levels of intentionality (1971). (For a note on the use of footnotes and inline citations in this thesis see the Reading Guide below.)

complexity and *limits* of the human capacity for processing intentional states, rather than its economy and expediency. In other words, the dominant questions seem to have been: "How many levels can X process?", or "How complex is behaviour Y in terms of the orders of intentionality?", rather than: "How can X perform behaviour Y with the least amount of processing power invested?"

(3) Various discourse topics inevitably involve embedded intentional states (such as the above-mentioned organisation of a dinner party, gossip about a second-removed acquaintance, and surely double spies and practical jokes). How are they handled in actual language usage? What can be learned from the way in which such topics are represented in natural discourse examples—that is: examples not invented in the context of philosophical analysis or psychological experiments, but taken from novels, plays, journalistic discourse, and spoken language?

(4) The conceptualisation of complexity as implying a series of embedded layers was *implemented* in various experimental paradigms and in (evolutionary) approaches to phenomena ranging from moral behaviour and cooperation to understanding and producing language and literature. In view of possible alternative measures of complexity (as suggested in (1)) and analyses of actual discourse examples (as suggested in (3)), how can such implementations be evaluated? What are the consequences for such implementations when we start from a focus on the economy of processing intentional states, rather than on its limits (as proposed in (2))? On this basis, which recommendations can be made for future theoretical and experimental research?

The overall purpose of this thesis is to work out these observations and questions in detail, and in doing so, to rethink the nature of the complexity posed by networks of multiple, mutually connected and interlinked intentional states. Ultimately, I aim to contribute to an alternative view on how we handle such networks *linguistically*, in discourse, how their nature should be construed *conceptually*, and how we manage to process them *cognitively* without undue

strain. This brings us to the question of the title of this thesis: why is it called *The Lazy Mindreader*? The long answer clearly is in the chapters that follow, but the short answer is this: we humans live in a socio-cultural environment that allows us to be "lazy" regarding the investment of mindreading efforts most of the time. This environment, of which the *conventions* underlying language and interaction are an important part, contains the coagulated experience of many generations interacting with each other and the world around them. Globally, I think that most approaches to mindreading have placed too much of a burden on individuals as "isolated cognitive units", and paid too little attention to the ways in which this burden can be alleviated by, for instance, lexical items, grammatical patterns, or narrative strategies, and by the interlocutors (including writers/narrators) we cooperate with to make interaction work. My alternative view focuses on economy and least effort: processing of complex networks of intentional states is not seen as something the lazy mindreaders envisaged in this thesis do by default, but rather as a skill that is needed when the context requires *deviation* from a default—and even then, I will argue, is these mindreaders' processing often supported by mechanisms that are part of, mediated by, or closely tied to language and narrative.

Structure

The body of this thesis consists of six chapters, four of which read as independent studies (see the Reading Guide below for details). The basic framework will be laid out in Chapter 1, where the concept of intentionality is discussed in detail and positioned in the broader literature on "mindreading", the capacity to assess intentional states of others and oneself. This skill, also known as "theory of mind", "mentalising", or "folk psychology", has been studied extensively in a multitude of academic disciplines across the humanities and sciences. I will distinguish three different ways in which the relationship between language and mindreading has been construed throughout the literature, namely: language as a way to represent mindstates and their mutual relationships, language as a "device" providing support to (the development of) the profound mindreading skills typical of human adults, and language as being itself crucially supported by our mindreading capabilities.

Finally, the role of mindreading in the bigger story of the "social brain hypothesis" will be assessed. Many of the concepts and terms discussed in Chapter 1 will be built upon throughout the rest of the thesis.

In Chapter 2 the focus will centre on the question of how literary texts represent complex networks of intentional states. Shakespeare's *Othello* will serve as a case study. Already by the end of the play's second act, a reader or watcher will understand that Iago *intends* that Cassio *believes* that Desdemona *intends* that Othello *considers* Cassio's rehabilitation. While this proposition may look as opaque and hilarious as Phoebe's most-renowned sitcom quote cited above, it is also in some sense a fair representation of (what I will call) the "thoughtscape" that has emerged at this point of the play's plot. Since there is no doubt that this plot has been understood and appreciated by many different audiences for ages, regardless of whether they were reading the text or watching a performance on stage, the question can be put as follows: what did the play do to make this thoughtscape manageable and accessible without undue cognitive strain? A detailed answer will be provided by distinguishing six *expository strategies* available in narrative discourse, which support the audience in gradually developing a robust understanding of the complex network of embedded viewpoints involved in the plot. These strategies are: characterisation, focalisation/viewpoint management, framing, episodic structuring, time management, and redundancy.

Chapter 3 takes this discussion into the realm of the novel. Cognitive literary scholar Lisa Zunshine has argued that in order to understand and appreciate a work such as Virginia Woolf's *Mrs Dalloway*, readers need to complete strands of reasoning of the following form: "Woolf *intends us to recognize* [...] that Richard *is aware* that Hugh *wants* Lady Bruton and Richard to *think* that because the makers of the pen *believe* that it will never wear out, the editor of the *Times* will *respect* and publish the ideas recorded by this pen" (Zunshine, 2006: 33, italics in original). This is again a proposition that is virtually impossible to process correctly by itself. I will demonstrate that novels, and in particular Modernistic works such as *Mrs Dalloway*, provide promising material for studying how multiple intentional states (in the context of literary analysis more generally referred to as *perspectives*) can be construed, entertained in series or in parallel, and mutually coordinated using linguistic

and narrative techniques. One finding will be that the conceptualisation of multiple-order intentionality as a series of embedded layers yields an unnatural and artificial fit with the actual material. The concept of *polyphony*, adapted from the work of Bakhtin (1984), will be introduced and discussed as the basis for an alternative conceptualisation: instead of focussing on complexity through embedding only, it outlines a model in which complexity consists of mental states being mutually related and interlinked in all kinds of different ways, delivered to the reader in manageable "chunks" by structural features of the text.

In Chapter 4, a shift will be made away from literary and narrative analysis to linguistics, with a view to examining the grammatical and semantic phenomena involved in handling multiple intentional states in journalistic discourse. In order to explore the nature and range of these phenomena, an analysis will be presented of newspapers' reporting on the so-called "Pistorius case".[3] Right after the shooting, journalistic sources would never have printed statements of the type: "a spokesperson *stated* that police officials *declared* that the athlete *claimed* that he *thought* that he was shooting at a burglar, while the responsible police detectives *claimed* that he *knew* it was his girlfriend". However, if we think about it, that *is* more or less the content of what they wrote, but they "packaged" the perspectives underlying the case into more convenient expressions, such as "Oscar Pistorius *allegedly accidentally* shot dead his girlfriend". Special attention will be paid to the discussion of what will be termed "viewpoint packages", single lexical items implying one or more viewpoint layers, such as *alleged(ly), accidental(ly)*, or *mistaken(ly)*. Also, the suggestion will be developed that viewpoint packages not only serve efficient communication of situations involving multiple mindstates, but can also support cognition: they may function as "thinking tools" acquired in a socio-cultural environment, supporting and enhancing our ability to process multiple-order intentionality cognitively.

[3] On February 14th, 2013, Olympic athlete Oscar Pistorius, also known as the "Blade Runner" on account of his blade-like carbon prosthetic legs, shot dead his girlfriend Reeva Steenkamp in their home in Pretoria. While claiming that it was an accident, he was arrested on the charge of murder. In the aftermath of the shooting, news media all over the world reported on the incident, covering the perspectives of the athlete, police officials, witnesses, family members, and others involved.

Chapter 5 will point out a consistent pattern throughout the material discussed in the preceding chapters: whenever multiple intentional states are involved, the labour of representing and coordinating these in discourse is distributed over a variety of narrative features and linguistic elements across multiple lexical, grammatical, and narratological categories. I will demonstrate how the developed view provides an important link in resolving a much-debated issue: the extent to which we need to engage in complex mindreading tasks when using language. Some researchers, most notably Sperber (2000) and Scott-Phillips (2015), take it that linguistic communication requires processing of what the other *intends* one to *understand* that the other *wants* one to *believe* by using a particular expression. However, building on Clark (1996) and Verhagen (2015), I argue that interlocutors in a (linguistic) interaction event should not be seen as two distinct cognitive units using a combination of language and mindreading for "pairing" their individual intentional states, but instead as one joint cognitive unit using language and mindreading in order to negotiate how a set of shared beliefs (or *common ground*) should be updated. While the first conceptualisation presupposes heavy and complex mindreading by default, the latter suits the "lazy mindreader" much better: only in exceptional cases, such as repairing a misunderstanding, complex mindreading enters the picture. This argument will form the basis for suggesting an updated version of the "construal configuration" model as previously developed by Langacker (1990) and Verhagen (2005). Using the updated model and its graphic representation, features of single linguistic elements can be highlighted along three axes, corresponding to three different types negotiation between a speaker and addressee about how to update the common ground. The y-axis indicates features that negotiate a relationship between interlocutors and objects that they jointly attend to, for example: "that football player". The x-axis is for indicating negotiation of epistemic stances between interlocutors, such as "that *great* football player". The z-axis deals with the negotiation of the degree to which objects of joint attention are considered from the perspectives of third-party discourse participants, as in "that *so-called* great football player". Here, "so-called" implies that another party, not specified here, is responsible for the

qualification of the person attended to as a great football player.[4] The merits of adding a third dimension to the existing model of the construal configuration will be discussed in the light of various examples and existing approaches to intersubjectivity, viewpoint management, and epistemic stance marking.

Chapter 6 is concerned with the practice of assessing multiple-order intentionality experimentally. In studies part of what I refer to as the "mentalising paradigm" participants were asked to read, listen to, or watch short stories describing a particular sequence of social interactions, for instance: the organisation of a surprise party. These stories were then followed by questions of the form "Did A know that B wanted C to come to his party?", "Did C know about the party?", or "Did B want A to think that C should know about the party?". Scores on such tests are being used as indicators of the "mentalising capability" of each individual participant, and have been shown to be associated with various assessments of people's real-life social functioning, such as the size of their social network. However, as yet little is known of the mechanisms and cognitive functions these tests actually tap into, and consequently, of *how* precisely these associations must be explained and understood. I will make suggestions for a "reanalysis" of the questionnaires and the results that they have produced, based on insights accumulated in the preceding five chapters.

Methodology

Whereas the topic of this thesis requires the study of research done in a wide array of different disciplines across both the humanities and sciences, its methodology is clearly rooted in the humanities. In the first place this is reflected in the overall endeavour being *qualitative* and *analytical* in nature, rather than experimental, correlational, or computational. This does by no means entail that it is not *empirical*: insights will be developed and evaluated in

[4] In fact, it should be noted that "so-called" is a linguistic element operating not just along the z-axis, but also along the x-axis, since it simultaneously signals a negative epistemic stance of the speaker towards the football player: "that so-called great football player" could be paraphrased as "others say he is a great football player, but I would not say so". In the terms introduced in Chapter 4 and 5: "so-called" is a *viewpoint package* with a *topology* coordinating mindstates along the z- and x-axes.

constant interaction with examples taken from actual literary, journalistic, and spoken discourse, and by examining questionnaires and test results in use in current experimental paradigms. However, instead of being primarily interested in finding significant statistical associations or constructing apt computational models, the aim of this study is to come to a coherent and well-wrought conceptualisation of the problems and their possible solutions in the targeted domain, in this case: the nature of the complexity involved in dealing with multiple interconnected intentional states, as required by particular aspects of our social and cultural environments.

In terms of Marr's (1982) "levels of explanation", this thesis operates for a large part on what has been labelled the "W-level" of *what* a system (in this case our ability to handle multiple-order intentionality) does, *why* it does this, and under *which* conditions it operates. In some places it also operates on the "H-level" of *how* the system works and which mechanisms are involved. The "Ph-level" or "physical level", which questions how the system is realised physically, is discussed only to a minimal degree.[5] As such, the approach chosen in this thesis runs counter to the majority of research on this topic, which is mostly about the How-level and Physical level, but rarely addresses the What-level. Behind this choice of focus lies the conviction that in order to make progress on the levels of the mechanisms and their implementation in our cognitive structure, an accurate grasp on the nature of the task is necessary. The "output" of the analytical work done in this thesis aims at providing such a grasp, and will ideally inspire new rounds of experimental testing, formal modelling, and neuroscientific investigation by, or in collaboration with, researchers who have their backgrounds in such sciences.

The "humanities roots" of this thesis are not only reflected in the qualitative and analytical nature of the approach, but also in the choice of concepts and methodological frameworks. The analysis of the literary texts, excerpts, and examples from journalistic and spoken discourse builds on narratological theory and on what is generally referred to as cognitive linguistics, cognitive literary studies, and cognitive stylistics. In Chapters 2, 3, and 4, there will be an important role for Dancygier's (2012) "narrative spaces

[5] Geurts and Rubio-Fernández (2015) have introduced the terms "W-level" and "H-level" for what Marr (1982) refers to as the "computational" and "algorithmic/representational" levels.

framework", which itself makes extensive use of the concepts of mental spaces, blending, and framing. Moreover, Chapter 5 will build on Clark's (1996) notion of "common ground" and Verhagen's (2005) "intersubjectivity" and his model of the "construal configuration". At the same time, research from various disciplines across the social, biological, and cognitive sciences will be used. In Chapter 1, Dennett's (1983; 1987) work on intentionality, which itself is based on discussions of ethology and animal cognition from the 1960s and 1970s, will play a central role. In addition, I will build on Apperly's monograph *Mindreaders* (2011) and include the wide-ranging research done by Dunbar and colleagues in the context of the social brain hypothesis.

All in all, the approach presented in this thesis is thus multidisciplinary in two ways.[6] Firstly, it combines and integrates insights and analytical tools that have a long history within multiple disciplines from the humanities: literary studies, linguistics, and philosophy. Secondly, it seeks to contribute to a debate that has pervaded research in diverse fields and traditions across both sciences and humanities, including psychology, cognitive neuroscience, ethology, philosophy, and cognitively-oriented literary studies and linguistics. As such, this thesis embodies a plea for what I call *topic-oriented scholarship*: it takes a topic as its starting point and then seeks for the right combination of methods and expertise across multiple disciplines for approaching it, instead of starting from the set of questions and assumptions customary in a particular discipline. Thereby, it aims at making progress not just by contesting existing findings, but also by adding new perspectives on these findings. Hopefully, these perspectives will inspire researchers from both the sciences and the humanities in their future, ideally joint, research on this topic.

[6] In my view, "multidisciplinary" is the appropriate label when research done within multiple disciplines is brought together; "interdisciplinary" means that methods and conceptual frameworks from one discipline are applied in another one. In that sense, this thesis as a whole is a multidisciplinary project, but it also takes an interdisciplinary approach at times: for example, in Chapter 6 I apply linguistic and narrative analysis to tests used in experimental psychology.

Reading guide

The format of this thesis can best be characterised as "hybrid": it holds middle ground between a classic thesis in book form, as has long been common in the humanities, and a collected set of related though independent papers, as is common throughout the natural and social sciences. As laid out in the Introduction, its core consists of six chapters. Chapter 1 and the Conclusion are written mostly in service of the other chapters, introducing the broader context, concepts, methods, and terms, and eventually wrapping up and tying all lines together. Whereas all chapters in principle use inline citations, Chapter 1 features footnotes wherever the introductory nature of the text would otherwise be jeopardised by too long lists of citations.

Chapters 2, 3, and 4 read as independent studies. They begin with an abstract and have a paper format, rather than a chapter one. Chapter 2 was published separately in *Langauge and Literature* (see Van Duijn, Sluiter, and Verhagen, 2015) and inserted without modifications. Paper versions of Chapters 3 and 4 are currently under review at a literary journal and a journal with a focus on cognitive linguistics respectively. Parts of Chapter 5 (especially Sections 5.3 and 5.4) have been presented at the Societas Linguistica Europaea (SLE) conference, taking place from 11-14 September 2014 in Poznan, Poland, and at the Perspective Project Kick-off Meeting, on 17 November 2014 in Nijmegen, The Netherlands. A paper version of this chapter will be submitted for the next SLE volume (forthcoming 2016). Chapter 6, then, contains the most practical part of this thesis: it offers a detailed analysis of selected stimuli and questions from three studies done within the mentalising paradigm, thereby aiming at "exporting" the insights of the other chapters to the practice of the lab, hopefully inspiring future rounds of experimental testing.

The chosen hybrid format has pros and cons. To start with a disadvantage that is particularly manifest when reading the thesis from cover to cover: some (especially introductory) parts of the chapters are repetitive. However, it was neither possible nor desirable to eliminate such repetition, given that each separate study interacts with different academic fields and bodies of literature,

and targets different audiences (roughly: psychologists and cognitively-oriented literary scholars and linguists in Chapter 2; literary scholars in Chapter 3; linguists in Chapter 4; psychologists, linguists, and philosophers of language in Chapter 5; philosophers and experimental psychologists in Chapter 6). Therefore, although certain portions may at times overlap, the focus is different in each chapter, and so are the choices of what to highlight and what to take for granted.

Despite the inconvenience this may cause anyone reading the entire thesis, in the long run I believe the chosen mode of presentation can yield an important advantage: it will hopefully enable the separate chapters to each independently find their way into the different fields and reach the different audiences for which they were intended, without the "ballast" of being available only as the part of a whole thesis written to fit one particular field.

Chapter 1

The bigger picture: language, narrative, and social cognition

Any mentioning of *intentionality* or *intentional states* comes with an interesting assumption: that we can speak meaningfully about the inner lives of others and ourselves. Indeed, everyday language is packed with "mentalistic" expressions of the type: "I *know* what you're after", "he *thinks* she's married", "she *believed* that he *feared* nothing more than that", etcetera, and in numerous contexts it is perfectly unproblematic to speak and reason in this way. At the same time, shifting from a run-of-the-mill perspective to one of philosophical and scientific inquiry, one may legitimately ask: what *do* we know about someone else's beliefs, thoughts, intentions, desires, fears, and so on? And what ways do we have available to form an understanding of this? A different question may be: why would we bother at all?

These questions have been central to the research into *social cognition*, the sort of cognition required for living in groups structured by social bonds and networks. As mentioned in the Introduction, researchers from a wide array of disciplines have contributed to this area, most notably psychologists, philosophers, anthropologists, ethologists, and neuroscientists. Most attention has been focused on the skill referred to as "mindreading" (also variously called "theory of mind", "mentalising", or sometimes "folk psychology"; see below), the capability to assess others' intentions, knowledge states, motives, etcetera—in short: their intentional states. As also set out in the Introduction, the main endeavour of this thesis consists in analysing the nature of the complexity involved in dealing with multiple intentional states that are mutually linked and/or embedded, as required by various aspects of our social and cultural lives, and investigating ways in which we handle such complexity linguistically and cognitively. Evidently, properly addressing the issue of handling *multiple* intentional states first requires knowledge of what it entails to form an understanding of just *one* intentional state. This chapter will start with a discussion of several possible views on this question, as given in the wider area

of research into mindreading. Next, a brief status quaestionis of research into multiple-order intentionality will be provided. The chapter will end by considering the links between, on the one hand, mindreading and multiple-order intentionality, and, on the other hand, language, narrative, and various aspects of social interaction more widely, as discussed throughout the literature on (primate) sociality and social cognition.

1.1 Mindreading and intentionality

1.1.1 Mindreading

Research into mindreading easily attracts attention, though not always for the right reasons. It all too often conjures up associations with myths, fairy tales, science-fiction stories, or even with fortune tellers and crystal gazers of the suspicious sort. In modern science there is of course a complete consensus that there is no magic involved in the way we form understandings of others' inner lives—however, anyone trying to come to grips with the extensive literature on mindreading that has emerged over the past decades might well form the suspicion that this is indeed the *only* consensus. To give a (rough and preliminary) impression: some research traditions have pictured a dedicated mindreading "module", forming the quintessence of the human mind (see e.g. Saxe, 2006). Others, by contrast, have conceived of mindreading rather as an "umbrella term" for a set of diverse tricks, strategies, and mechanisms that we use to make sense of the behaviour of ourselves and others around us (e.g. Apperly, 2011). Some have emphasised the role of brain functions specialised for mindreading (e.g. Carruthers, 2004), others have suggested that we use only general cognitive skills (e.g. Heyes, 2014). Some are particularly interested in the aspects of mindreading that are uniquely human, others emphasise their deep roots in our primate (or even mammalian) nervous systems (e.g. De Waal, 2013). Some maintain that mindreading relies on innate competencies (e.g. Fodor, 1983), whereas others stress that the most important parts of mindreading are learned in the course of growing up in our typically human socio-cultural

environments (e.g. Heyes and Frith, 2014). According to some researchers, mindreading is highly "enactive" and performed by using our entire body for modelling someone else's perspective (e.g. Gallagher, 2008), while others have suggested that we run simulations of what people around us feel and think using "mirror systems" in our brains (e.g. Gallese and Goldman, 1998). At the same time, defenders of an "inferentialist" understanding of cognition have suggested that we form representations of and theories about the inner lives of others instead of running simulations of any form (e.g. Gopnik and Wellman, 2012). Advocates of the "narrative practice hypothesis", in turn, have argued that mindreading most often relies neither on simulation nor theorising, but rather on structural and semantic knowledge of folk-psychological narratives (e.g. Hutto, 2008). Several of these positions will be detailed and built on below.

At the outset of his monograph *Mindreaders* (2011), Apperly provides a comprehensive overview of the main questions and debates that have occupied researchers of mindreading over the past decades. His book focuses mostly on explaining how mindreading works in terms of its underlying mechanisms, which means that it operates for the largest part on Marr's H-level ("how", "through which mechanisms"; see the Introduction). In line with the purposes of this thesis, the discussion in this chapter is mainly focused on the W-level ("what"): it aims at setting out a workable "task model" of mindreading by discussing the elements and stages of its process and the conditions under which it operates. Nonetheless, this chapter also contains several sections pertaining to the mechanistic and physical levels of explanation, as parts of introducing the larger field of research.

Stripped down to its basic outlines, the task model set out here features five elements:

(i) the mindreader;
(ii) the mindreadee;
(iii) cues;
(iv) intentional states (which can be called "mindreads" once the mindreading process has taken place); and
(v) an inferential process through which (iv) is derived from (iii)

I will begin by discussing an introductory example, exploring some of the issues and terms that will be revised and built upon in the sections that follow. Consider the following photograph:

Figure 1

A normally developed adult person standing in the position of the photographer (the mindreader) will most likely feel inclined to give the person on the staircase (the mindreadee) a helping hand. The term used for the basis of this inclination by Frans de Waal (2005) is *emotional contagion*:[7] we see the facial expression and posture of the person carrying out a heavy task and due to the deeply-rooted empathic tendencies we have as primates, we cannot even help but feel some of the burden ourselves, which triggers the impulse of providing targeted help (more details and alternatives will be discussed below in Section

[7] "Emotional" is here to be understood not in the narrow sense of the "basic emotions", but rather as the broader category of feelings including, for instance, pain, grief, agitation, relief, sorrow, embarrassment, surprise, and so on (De Waal, 2005: 46-47).

1.1.4). Whether this help is in the end provided will clearly depend on many factors, such as individual features of the mindreader, relationship to the mindreadee, local cultural rules, and more. However, it is not hard to see that Figure 1 above depicts a situation in which a helping hand would *in principle* be appropriate. Now consider the following photograph:

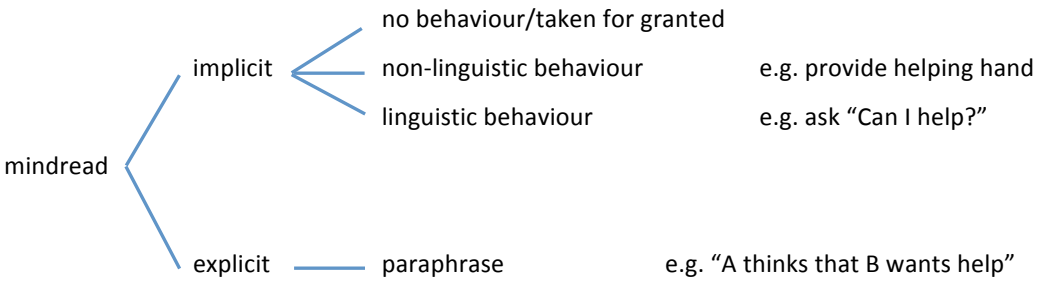
Figure 2

The weightlifter depicted in Figure 2 shows a posture and facial expression considerably similar to that of the person in Figure 1 and the tasks faced by both are also much alike: lifting a heavy object. Yet it is clear that this man would show much surprise, if not severe agitation, if the photographer or any of the other bystanders would offer a helping hand—and indeed, none of them shows the least inclination to come to his aid. (A caption making this point in a different way would be: "Why is this *not* an example of the bystander-effect?") Even if the weightlifter were to look up and ask for a helping hand, the mindreader would probably start mining the situation for opportunities to provide assistance with anything *other than* just lifting the weight bar, such as a

loose shoe lace or an unfastened spring collar at one of the bar ends—at least, that is, as long as the weightlifter gives the impression that he has the situation under control. This highlights another crucial component of the inferential process: *background knowledge*—not only emotional contagion influences the mindreader's decision on whether to take action or not, but also knowledge about the situation the mindreadee is in. In principle, one could imagine that someone completely unfamiliar with the context of a gym would hasten to help lifting the weight bar (some philosophers would suggest an empathic Martian; others would perhaps suggest an unworldly philosopher). He would pick up some of the burden felt by the mindreadee, while being unable to recruit the relevant background knowledge about what the possible *scenarios* are in this context. Someone who does know the context of a gym, by contrast, does have such scenarios available: the weightlifter wants to test or train his strength, or possibly show off to the bystanders. Clearly, in these scenarios help is highly unwanted. If he were lifting the bar from his car boot, though, a helping hand might again fit.

The decision about which behaviour is appropriate in the situations depicted by Figures 1 and 2 relies on what I here call a *mindread*: the assessment made of someone's intentional state in the context of a (real or imagined) social interaction event. Such an assessment is made on the basis of *cues*, which can be of virtually any nature. What does or does not count as a cue can only be defined from the perspective of the mindreader: it includes any observable aspect exhibited by the mindreadee and his or her "situatedness" in the context of the interaction that is used in the mindreader's inferential process. The cues are interpreted in the light of relevant background knowledge, recruited from the mindreader's memory. In practice, the resulting outcome, i.e. the mindread, can be made explicit or remain implicit, and can be taken for granted or factored into the planning of future behaviour. This behaviour can be linguistic (the mindread can guide form and content of an utterance or response) or non-linguistic, as would be the case when providing a helping hand to the person in Figure 1. Moreover, for purposes of analysis or reflection it is possible to form explicit paraphrases of a mindread, for example: "this person *intends* to lift the suitcase in order to get upstairs, so *will appreciate* a helping hand" or "A *thinks* that B *wants* help" in the case of Figure 1. As will be discussed in Section 1.1.3

below and in Chapter 2, it is important not to confuse such paraphrases with the cognitive processing required to make appropriate inferences, and with representations of intentional states as they appear "in the wild" of actual discourse. The distinctions now made can be summarised as follows:



Sections 1.1.2 and 1.1.3 will further elaborate on element (iv) of the list above: intentional states. In Section 1.1.4 element (v), the inferential process, will be discussed in more detail.

1.1.2 Intentional states

The concept of *intentionality* (not to be confused with the "intentions" we have when we want something to happen) has a rich history in scholarship and sciences of the mind. After its presumed origin in medieval scholastics,[8] the concept was most famously developed in the nineteenth century by Franz Brentano (1995 [1874]), as a part of debates now considered foundational for the emergence of psychology as an academic discipline, and in the twentieth century by Daniel Dennett (1971; 1987) and John Searle (1983), in work that was influential in the still-ongoing trend in psychology and the cognitive sciences to study mindreading. Brentano used the concept of intentionality to define the difference between mental and physical phenomena. In brief, his distinction boils down to the claim that physical phenomena have an autonomous existence, whereas mental phenomena necessarily are *about* something—they do not exist independently of their intentional object. In Brentano's words: "in

[8] According Chisholm's (1967) entry in *The Encyclopedia of Philosophy* a very similar concept was already present in Saint Anselm of Canterbury's 1078 treatise on the existence of God, but the term was coined later and goes back to the scholastic notion of "intentionalitas".

presentation something is presented, in judgement something is affirmed or denied, in love loved, in hate hated, in desire desired and so on" (1995 [1874]: 88). This "something" that is presented, judged, loved, etcetera, is the intentional object on which each intentional state depends for its existence.

Making a distinction between mental and physical phenomena in this way has implications for how the ontology of mental phenomena is construed. Therefore, work on intentionality has always been connected to fundamental philosophical debates regarding "dualism" and, more recently, the mind-body distinction and extendedness of cognitive processes. Brentano seems to accept a meaningful divide between the physical or material world, in which objects exist *as objects*, and the mental world, which includes non-material phenomena that are *about objects*.[9] From a non-dualist, "materialist" viewpoint this position is problematic, since mental phenomena are being viewed as a part of the physical and material world in no other sense of the words. Within the materialist view, a distinction can again be made between, on the one hand, theorists who conceive of cognitive processes as neurons in the brain "dancing" in a particular way, and, on the other hand, theorists who argue that cognition is mostly distributed over the entire body, or even beyond that: over the environment. Defenders of this latter position, which is known as "extended cognition", allow for combinations of, for instance, people, machines, books, and/or cultural practices to be included in their definitions of cognitive processes. Some more attention to the issue of embodied cognition will be paid in Chapter 2, where mindreading is considered in the context of drama and dialogue. In Chapters 4 and 5, I will argue that part of the burden of processing

[9] Note that this does not mean that he argues that physical phenomena can be experienced unmediated by our senses: he sees the world as being entirely mediated by subjective experience; however, some parts of our experience relate to physical phenomena and others to psychological phenomena (see Zahavi, 1992: 30). In this sense, Brentano can be called a dualist, but not a Cartesian. See also Dennett (1987: chapter 10).

complicated mindreading tasks can be alleviated by cultural and linguistic "thinking tools".[10]

Dennett, who is an explicit defender of cognitive materialism, bypasses much of this controversy by speaking of "intentional systems" (1971; 1983). Whether a cognitive system is construed as an immaterial mind, a group of interconnected neurons "dancing" in a particular way, an entire body, or as two people and a calculator, in all cases it can be seen as a *system capable of intentionality*, in the sense that it can enter a state in which it is *about* an object. This object should be taken in the broadest sense of the word, including, for instance, a ball in the mouth of a dog, a comic figure on a computer screen, a picnic in a short story by a price-winning writer, or the creator of the universe. Whether the object exists in some form outside the realm of the intentional state is, in this view, not part of the question: intentionality is seen as a property that a system can have, regardless of how this system is realised and regardless of the ontological status of the object. In this way, and as mentioned in the Introduction above, an intentional system can enter a state in which it is about *another* intentional system, potentially also exhibiting an intentional state. For example, a human mind can be about another human mind's intentional state, say, the other's desire for a glass of water, intention to cooperate, or understanding of her brother's love for his daughter. Such embedded or multiple-order intentionality will be discussed further in Section 1.2 below.

Characteristic of intentional states, then, is that they have a "dependent" or "extending" nature: when considering an intentional state, one necessarily also has to take into account the object this state is about. A traditional way of studying this is using "logical" propositions. Consider the following expressions:

(1) John believes that it is raining outside.

[10] See Dennett (1987: chapter 10) for a taxonomy and discussion of various theories about intentionality and their implications for the ontology of mental phenomena. Although explicit defence of a classical "Cartesian" dualist view is rare in modern philosophy and science, implicit assumptions referring to this view can be traced in many works and research traditions (see also Dennett, 1991: chapter 2 and 5). The latter point is also demonstrated in Sorem's (2010) discussion of Searle (1992): Sorem argues how Searle's "simple solution" (1992: 1) to the mind-body problem yields inconsistencies, in a way that is illustrative of how deeply rooted dualistic views are in everyday thinking as well as in specialised philosophy.

(2) Mary intends that John believes that it is raining outside.
(3) Mary talks to John and it is raining outside.
(4) It is raining outside and Mary intends that John has another cup of tea.

They can be rewritten into propositions as follows:

(5) A believes that p
(6) B intends that A believes that p
(7) p and q
(8) p and Mary intends that q

Note that clauses expressing intentional states are rewritten using the form "A [intentional expression] that p", whereas other clauses, referring to objects (in the broad sense, so including events, states of affaires, people, etc.), are rendered as single symbols (p or q). This reflects the structural property of intentional states discussed above: they are not independent, but reflect a *relationship* between an intentional being and a non-intentional object. In logical terms: a clause expressing an intentional state induces "referential opacity", as can be shown by the so-called "substitution test" (Dennett 1983: 344-345). In a proposition describing a particular state of affairs in the world, it is usually possible to substitute words with other words that refer to the same entity without consequences for the truth-value (or even referential value) of the sentence. In other words: "this rule is simply the logical codification of the maxim that a rose by any other name would smell as sweet" (Dennett, 1983: 344). To give an example: provided that *Macbeth* and *Hamlet* were written by the same author, it should be possible to substitute "the author of *Macbeth*" with "the author of *Hamlet*" in a proposition, without the truth-value and referential value being affected. The propositions (9) and (10) are thus either both false or both true:

(9) The author of Macbeth was born in Stratford-upon-Avon
(10) The author of Hamlet was born in Stratford-upon-Avon

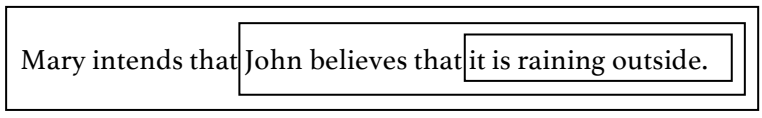
However, in the following two propositions this is not necessarily the case:

(11) John believes that the author of Macbeth was born in Stratford-upon-Avon

(12) John believes that the author of Hamlet was born in Stratford-upon-Avon

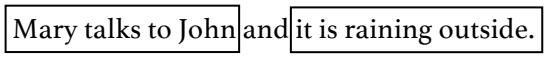
After all, what John does and does not believe is independent of the "real-world fact" that the author of both pieces is the same person. This is what Dennett calls referential opacity: "the terms in such clauses are shielded or insulated by a barrier to logical analysis, which normally "sees through" the terms to the world the terms are about" (1983: 345).[11]

The take-home message from the substitution test is really that there is a relation of dependency between the intentional agent and the non-intentional proposition. In the case of "John believes that it is raining outside", the intentional agent (John) and the intentional expression (believes that) are the responsibility of the speaker, the one who asserts the proposition, whereas the non-intentional proposition (it is raining outside) is placed under the responsibility of the staged intentional agent (John). As a consequence, it is "insulated", in Dennett's words, "shielded from logical analysis" (1983: 345). In "Mary intends that John believes that it is raining outside", there are two such dependency relations: "John believes that it is raining outside" falls as a whole under the scope of "Mary intends that", and "it is raining outside" falls under "John believes that". Schematically:

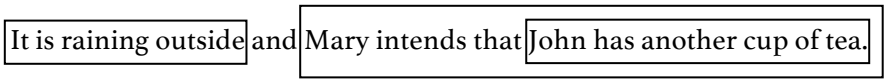


[11] There is a significant difference between this "logical" approach and the natural-language view mostly taken throughout the rest of this thesis. Instead of looking at references to the "real world", in this latter view the focus is on a speaker having a certain rhetorical goal (in this case presumably informing his interlocutor that John believes that the author of a particular piece was born in Stratford-upon-Avon). With a view to achieving this goal the speaker invites his interlocutor to consider the perspective of John, which he does (in accordance with local linguistic conventions) through the usage of a complementation construction (see Chapters 4 and 5 in particular).

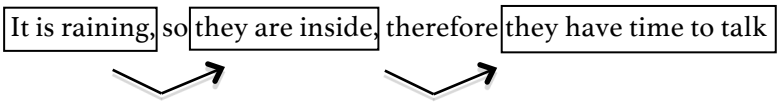
By contrast, in the case of a conjunction of two non-intentional propositions, there is no dependency in this sense:



Clearly, when an intentionality proposition is combined with a non-intentionality proposition, the latter is independent of the first:



Another aspect of intentional relations is what could be called their *non-transitivity*.[12] The following proposition features three clauses that are causally related:



In principle, it is possible to leave out the middle clause without violating the chain of causality expressed by this proposition:

In other words, if a proposition expresses "p so q therefore z", it follows that "p therefore z" is also true—causal relationships could therefore be called "transitive". By contrast, if one clause is left out of a chain of intentionally related clauses, it does not follow that the produced clause has the same truth-value: "B believes that A believes that p" does not entail that "B believes that p". The points made in this section will be of practical use when analysing questions from mentalising experiments in Chapter 6. Their theoretical importance will become clearer in Section 1.2 below and in the Chapters 2 and

[12] In formal logic transitivity is a property of certain relational predicates, such as ancestry. If A is an ancestor of B, and B is an ancestor of C, then it follows that A is also an ancestor of C (see e.g. Forbes, 1994: 275).

3; in fact, it can be said that the non-transitive, dependent nature of the relationships exhibited by multiple-order intentionality propositions is one of the core aspects of the problem dealt with in this thesis.

1.1.3 The intentional stance

So far in this thesis, different intentional relationships have been categorised using mentalistic expressions from everyday language, such as thinking, knowing, believing, desiring, intending, and so on. How can we be sure that these terms are appropriate? Do they correspond to the *actual* intentional states held by others around us? Or, for that matter, to those held by non-human animals? According to Dennett (1983; 1987) we do not need to be sure. He argues that in order to understand phenomena in the world, one can adopt various strategies or "stances", corresponding to different levels of theorising (partly overlapping with Marr's levels of explanation, as will be pointed out below). For example, for understanding why an analogue alarm clock rings, one could take the "physical stance" and aim at figuring out how, given the laws of physics, its springs exert particular forces on a system of cogwheels and axes, eventually triggering a clapper that hits a bell, which causes movement of air at particular frequencies, and so on. One could also take the "design stance", looking at what the clock was designed to do when setting it to a particular time and switching on the alarm function. Alternatively, one can adopt the "intentional stance", not towards the alarm clock itself, but to the intentional system who set it: one can question why someone has set it to this particular time and what he or she *intended* hearers of the alarm sound to *think*. If we decide to adopt the intentional stance, this means that we try to make sense of an intentional system's behaviour by assuming that it was underlain by intentional states. In this view, the intentional states are really "in the eye of the beholder", and their use is to understand a phenomenon in the world better.

Intentional systems can be humans, but also other animals, for example: "that fox digs a hole because it *wants* to build a nest" or "bird X *believes* that bird Y is hiding food". Dennett argues that usage of everyday language is not problematic in such cases, as long as one keeps to the appropriate level of theorising. To use an adapted version of his own example: when researchers

interested in the behaviour of a particular bird decide to call a certain berry in the bird's environment "food", they abstract from all kinds of biological and chemical details of nutrition and digestion. Biologists interested in such details might choose to refer to the same berry in terms of its composing sugars, acids, proteins, etcetera. Even if the latter research were still in an early stage and little would be known about the biochemical details of nutrition from this berry, those interested in the foraging behaviour of the bird could safely refer to it as "food" in their theories. Similarly, one can perfectly well make use of everyday mentalistic vocabulary as long as one is dealing with questions of some beings' behaviour in their social environments, and not with the "lower-level" mechanisms and physical processes underlying social living.[13]

In some ways this is reminiscent of the distinction made between "explicit" and "implicit" mindreads at the end of Section 1.1.1. Normally developed human adults can surely reason about mental concepts in an explicit way: they can think or talk about themselves or others in terms of their beliefs, desires, fears, etcetera, thereby explaining or motivating particular behavioural moves and choices. However, this does by no means entail that mindreading in the practice of social interaction constantly uses explicit mental concepts. Apperly's (2011: 3) analogy in the physical domain is the curve described by a ball thrown into the field: although we can in principle reason about angles, velocity, friction, and so on, in order to predict where it will land, this hardly reflects how we manage to make a catch in practice. The explicit mental concepts and linguistic paraphrases could be seen as the formulas dealing with angels, velocity, and friction: although they can in principle be used to form a mindread, they hardly reflect our reasoning in most cases of everyday interaction (more on mindreading and linguistic explications will follow in Section 1.2.2).

Note that Marr's levels of explanation can again do useful work in this context: both adopting the intentional stance ("the fox *wants*...") and engaging in explicit mindreading ("A *thinks* that B *intends*...") are unproblematic on

[13] See Dennett (1983: 344) for his version of this example. Related to this is the debate over the question whether the everyday mentalistic terms should be seen as temporary placeholders, used only until their "real" neurological correlates are figured out, or whether they have a different status. See Dennett (1987: ch. 10) for a discussion of his own perspective and various alternatives.

Marr's W-level, the level of "what the task is". Terms such as thinking, knowing, desiring, and so on, work fine when used to describe what a mindreading process is about and why it is taking place. However, this may change as soon as one is dealing with questions on Marr's H-level of the mechanisms at work or the physical level describing what machinery is used: it is likely that intentional terms such as thinking, knowing, or desiring have no role in, say, what drives a fox to dig a hole or what makes a bodybuilder lift a heavy weight bar on the physical level.

In the next section, I will offer a brief discussion of various hypotheses operating at the level of how the inferential processes underlying mindreading are carried out in practice, thus switching to the H- and physical levels.

1.1.4 The inferential process: theory, simulation, narrative practice

The concept of "theory of mind" goes back to discussions about the cognitive features and limitations of our close primate relatives in the 1970s. The foundational paper on this concept was published by Premack and Woodruff in 1978. They discussed experiments in which a chimpanzee, Sarah, was shown videotaped attempts of a human actor to solve particular problems, such as trying to grab a banana that was placed out of reach. Sarah (who, incidentally, was said to be familiar with the video screen from watching "commercial television") had little trouble matching the videotaped problems with photographs that pictured the "right" solution out of several options, such as using a long stick to bring the banana closer. Interestingly, it seemed to matter to her that the actor in the videos was her favourite trainer, for whom she clearly felt affection, since when the same part was played by another, less well-known acquaintance of hers in subsequent test rounds, she quite consistently chose photographs showing "bad" alternative solutions leading to "untoward outcomes" (Premack and Woodruff, 1978: 521).

It is worthwhile looking into this early paper in some detail, as it already addresses a few important issues that have since been discussed extensively in the literature on mindreading, and are mostly still under debate. After establishing that the results could not be explained by mere "physical matching" of objects, the authors discuss three possible explanations of how

the chimpanzee could have managed to match the videotaped problems to the right solutions: "associationism", "theory of mind", or "empathy". According to the first explanation, she would have solved the problem on the basis of familiarity with relevant sequences of action; in the authors' words: "when shown a sequence that one recognizes, but that is incomplete, one chooses the element that has the effect of completing the sequence" (1978: 516). The authors point out that Sarah most likely had similar experiences from her own daily life in the lab, but that she was not familiar with the exact sequences of action used in the experiment and that there was sufficient reason to assume that the presented problems contained at least some novel elements for her. They therefore grant the "associationism" explanation some credibility, but regard it insufficient to account for the totality of their findings.

The second explanation, "theory of mind", is described by Premack and Woodruff as follows: "In looking at the videotape, [the chimpanzee] imputes at least two states of mind to the human actor, namely, intention or purpose on the one hand, and knowledge or belief on the other" (1978: 518). They thus suggest that, according to this explanation, Sarah somehow went through the following strand of reasoning:

(i) "the human actor *wants* the banana and is struggling to reach it";
(ii) "the actor *knows* how to attain the banana"; and
(iii) that will lead to the situation depicted in photograph X rather than photograph Y (1978: 518; italics added).

The authors consider the third explanation, "empathy", to be identical as far as step (i) is concerned, but different for step (ii) and (iii). After imputing to the actor the intention to grab the banana (step (i)), according to the "empathy" explanation Sarah would put herself "in the place of the actor" (1978: 518) and choose the alternative consistent with what she would do in that situation.

Given that it mattered to Sarah's choices whether she saw her favourite trainer or a more removed acquaintance on the video screen, Premack and Woodruff favour the "theory of mind" explanation over the "empathy" one: after all, if she would picture *herself* in the presented situations, the actor's identity should not matter. They argue that this does not exclude "associationism" to play a role as well, and conclude as follows:

> in highly familiar situations, one's expectancies are based on existing associations. [...] In novel situations, however, one's expectancies are generated, we think, from theories, and are not the product of associative generalization. [...] There may also be developmental and inter-species differences in this regard. Young children and lower species may form expectancies by associative mechanisms, the former having yet to build any theories and the latter probably unable to build them; whereas adults and higher species may largely generate them from theory. (1978: 518)

Extensive debates followed this early discussion, partly revolving around questions of which primate species had such theory-of-mind abilities and to what extent, and for another part focusing on analogous competences in humans. Later on, important contributions were made by developmental studies and research addressing certain psychopathological disorders. It seemed that some people suffering from disorders along the autistic spectrum were well-described as having impaired theory-of-mind abilities, which was generally taken as a strong indication that there must indeed be some part or network in the human brain responsible for theory of mind (after all, "if it can break, it must be there"). Another boost to the field was given by the advances made in the neurosciences during the 1990s and 2000s, including also the discovery of mirror neurons, neural networks that are involved both when an action is performed and when this same action is observed in someone else.[14]

Notwithstanding the importance of all these subsequent findings and contributions to the debate, support for all three explanations given by Premack and Woodruff for how the inferential process works persists to the present day in one form or another. The aim in the remainder of this section is to offer a typology of the dominant positions in the current field. By way of illustration, work of proponents of each of these positions will be referenced and discussed in brief, but these discussions must by no means be taken to be exhaustive. For more comprehensive overviews, providing more extensive lists

[14] See Di Pellegrino et al. (1992) for one of the initial papers on the discovery of mirror neurons; for a full discussion see Pineda (2009). For an overview of neuroscientific research into mindreading see Frith and Frith (2006). For an overview of literature on mindreading and psychopathology see Baron-Cohen et al. (2013); Bird and Viding (2014). For mindreading across the primate world see Byrne and Whiten (1988; 1997); Rosati and Hare (2010); Whiten (2013).

of references to advocates of the different positions see Apperly (2011: especially chapters 2 and 7), and see Hutto (2008: especially chapters 8 and 9) for a critical perspective.

Theory-theory

Proponents of what is referred to as the "theory-theory" hypothesis suggest that mindreading relies on folk-psychological theories, generally held to comprise knowledge of rules and patterns of how social beings behave, and why they behave like they do, with a certain degree of *abstraction*. So mindreading competence is hypothesised to rely not (just) on knowing a collection of sequences of action that can be used as "exemplars" when making sense of new cases, but rather on more abstract rules and patterns that can be used to generate predictions of future intentional states and/or behaviour. There are multiple versions of the theory-theory hypothesis in circulation, primarily differing in two dimensions: "specialised versus domain-general", and "innate versus learned". High scores on both specialisation and innateness can be associated with, for example, Fodor's work in this area. Roughly, his idea is that humans are born with "innately cognized propositional contents" (Hutto, 2008: 144, citing Fodor, 1983: 85), which can be understood as specialised modules containing the basic rules of folk psychology. In the practice of social interaction, these rules can be applied to representations of someone's beliefs, desires, and other intentional states, in order to yield predictions of someone's future intentional states or behaviour, much in the fashion of steps (i), (ii), and (iii) above, as suggested by Premack and Woodruff. According to this view, mindreading capabilities are in place from birth, but they are initially "masked": infants lack the ability to exploit their innate understandings of intentions, desires, beliefs, and so on, until they improve general skills such as selecting and processing information, and applying it appropriately (for a more recent defence of a position along these lines see Leslie, Friedman, and German, 2004).

Other researchers suggest that folk-psychological theories are constructed rather than inherited genetically. Most notably, Gopnik and Wellman (e.g. 2012) argue that children use "data" gained from experience with their own and

others' actions in the social world in order to infer causal structures using forms of statistical learning. In the course of growing up, they may test their theories via "informal experimentation" through play, and further refine them through imitation and pedagogy. According to this view, forming theories about others' inner lives and social behaviours is done using the same mechanisms as forming theories about other aspects of the world (such as behaviour of physical objects). Gopnik and Wellman are thus situated at the other end of the theory-theory spectrum: according to them, mindreading is learned using more domain-general mechanisms, instead of relying on innate and specialised structure.

Simulation theory

The second dominant view on mindreading is known as "simulation theory". The essence of this view is that the inner lives of others can be *modelled* using one's own mind. Like in Premack and Woodruff's early discussion of what they call "empathy"[15], the basic idea is that one reasons "as if being in the other's shoes". However, in more recent accounts of simulation theory a distinction is being made between, on the one hand, low-level simulation of actions, bodily expressions, and basic emotions, and, on the other hand, the high-level simulation of intentional states providing the motivations and conditions behind these actions, expressions, and emotions (see e.g. Gallese, 2001; Goldman, 2006; for a broader discussion and critique see Gallagher, 2012). Roughly, the opinion among simulation theorists is that the low-level component is present in infants and other primates, whereas the high-level is unique to humans and develops throughout childhood. The low-level component is argued to rely primarily on activation of mirror neurons,

[15] There are many different usages of the term "empathy" around in the literature on social cognition. Sometimes it is used as a synonym for mindreading, sometimes it is framed as a process underlying mindreading, and sometimes it is argued that the two can do without each other (see e.g. Gallagher, 2012, for a discussion of different positions among simulation theorists). In general, I think that the meaningful categories in the domain of mindreading and social cognition are covered in this section, though sometimes in a simplified form. It is possible to discuss some of these categories in light of the term "empathy", or even to re-label some categories with terms such as "affective empathy", "emotional empathy", or "cognitive empathy" (e.g. De Waal, 2005; Uzefovsky et al., 2015), but this is rather a different way of cutting the same cake than an extension yielding genuine conceptual enrichment.

providing a very direct sense of another's body movements, facial expressions, and (through that, probably) basic emotional states (see Note 7 above; see also Knoblich and Sebanz, 2006, for what they refer to as "common coding"). When given the example of the lady with the suitcase on the stairs discussed in 1.1.1 above, simulation theorists would probably explain the "contagious" effect of this scene in terms of low-level simulation processes making one "take over" part of the burden—after all, the mirror-neuron view predicts that some of the same networks in the brain are activated when *executing* the action of lifting the heavy suitcase as when *seeing* someone else do this. Simulation theorists would probably go on to add that the higher-level component of simulation is needed to distinguish between the situation on the stairs and the one in the gym: after all, only after running a full imaginative simulation of both situations ("putting oneself in the shoes" of the two mindreadees in the pictures), can one become aware that the lifting of the heavy object has a different goal in either case and thus is underpinned by a different set of motivations.

Narrative-practice and two-systems approaches

The third and fourth views that have come to prominence in the literature on mindreading are the "narrative-practice" hypothesis (Gallagher and Hutto, 2008) and the "two-systems" approach (Apperly and Butterfill, 2009; Apperly, 2011: chapter 6 and 7). These two views can be characterised by their attitudes towards the first two positions: defenders of the narrative-practice hypothesis tend to argue that *neither* theories nor simulations can account for how we understand the inner lives of others, whereas two-systems thinkers generally grant the importance of elements of *both* theory and simulation. However, this difference is one of emphasis rather than of essence, since upon a closer look the two approaches have much in common. Both start from the view that the foundations of social interaction lie in the mutual coordination of actions and body movements, allowing for such "embodied" routines as mimicry, alignment, and imitation, and both approaches suggest that these processes are predominantly automatic, in place from early infancy, to some extent shared with other primates, and that the mirror-neuron system might play an important role. In fact, their suggestions of how the foundations of social

interaction work, are much in line with the low-level part of simulation theory. However, it is characteristic of the narrative-practice approach to emphasise that social interaction at this level has nothing to do with assessing intentional states. According to Gallagher and Hutto, the most prominent adherents of this view, the entire understanding of others around us in terms of intentional states, is an *a posteriori* dimension which we apply to the social world and its inhabitants using our experience with folk-psychological narratives. It is only because children are *told* (and adults keep telling each other) what people think, intend, desire, etcetera, under which conditions they do this, and how this is linked to behaviour, that we make sense of interaction events by referring to "underlying" intentional states (see Gallagher and Hutto, 2008; Hutto, 2008; Gallagher, 2012). Although there is debate over the precise implementation of this view in the practice of actual social interaction, an important part relies on the "matching" of previously collected exemplars with the case at hand.[16] In this sense, this view comes closest to what Premack and Woodruff referred to as "associationism", with the important difference that chimpanzees of course only have their own experiences and observations available as exemplars, whereas the crucial advantage for humans is our access to culturally accumulated experiences and observations through narratives—we can gain experience through others' eyes, as it were.[17] A related though more specific suggestion regarding the acquisition of exemplar cases through language and narratives will be made in Chapter 4 and 5 of this thesis, and be integrated in the synthesis developed in the Conclusion.

Two-systems thinkers share with advocates of the narrative-practice hypothesis the view that narratives and social schemas are highly important for our understanding of the social world and the inner lives of its inhabitants, but narratives and schemas have a different place in their model. According to two-systems theory it is pivotal to recognise that social conventions have a high degree of normativity, and that the settings of (in)formal instruction through

[16] Incidentally, the building of *abstractions* on the basis of these exemplars is here not being excluded by the authors, which definitely blurs the sharp distinction with theory-theory they make elsewhere (e.g. Gallagher and Hutto, 2007).

[17] I discuss this view in more detail in Van Duijn (2015, in Dutch). A position that has much in common with this view, but does not put as much emphasis on narratives, can be found in Heyes (2012) and Heyes and Frith (2014). See also the summary of my position in this debate the Conclusion.

which children are familiarised with these conventions are aided by the practice of telling narratives (Apperly, 2011: chapter 6; see also Warneken and Tomasello, 2006). However, where narrative-practice thinkers have a tendency to downplay the importance of mindreading altogether, replacing it by non-representative forms of "bodily" coordination on the one level, and narrative competences on the other, the two-systems approach rather uses bodily interactions, narratives, social norms, and schemas to explain how mindreading at various levels is *possible*. The basic idea here is that, in order to serve the actual practice of social interaction, mindreading has to be quick and flexible at the same time (Apperly and Butterfill, 2009). If it were the case that for every word, gesture, coordinated movement, helping hand, etcetera, a full mindreading process had to run, based on the totality of cues, represented intentional states, knowledge of folk-psychological rules or simulation of the others' position, and so on, fluent interaction would be impossible. Therefore, two-systems theory suggests that part of the burden is taken away by quick, deeply-rooted, mostly automatic, bodily interaction routines, and that knowledge of social norms, schemas, and narratives can help a great deal in constraining the amount of information that has to be processed, and selecting what is relevant in a particular context, thus making the mindreading task tractable.[18] However, all of this may come at the cost of the flexibility needed when one is confronted with mindreading tasks that go beyond bodily routines, general schemas, and so on. Therefore, according to two-systems theory, there is at least one other system available on top of the basic, "quick" system, which can deal with non-straightforward cases in a more explicit, flexible, though slower and more cognitively demanding way, possibly using elements of both theorising and simulation (see Apperly and Butterfill, 2009; Apperly, 2011: chapter 6 and 7; see also Kahneman and Tversky, 1982).

[18] In fact, Chapter 2 works out this line of thinking for stories that appear to comprise highly complex mindreading tasks: I argue that a combination of "expository strategies" makes these tasks tractable, so that the audience in principle need not more than basic mindreading skills to be able to follow the plot.

Blind men and the elephant

By way of concluding this section, I will discuss a little thought experiment adapted from a study by Kahneman and Tversky (1982; also discussed by Gallese and Goldman, 1998) leading to a brief summary and synthesis. Imagine that two travellers share a taxi on their way to the ferry port. Each of them has to take a different ship, however, both ships are scheduled to depart at the same time. During the taxi ride they are confronted with unexpectedly heavy traffic and, on top of that, when they are nearly there, the driver takes the wrong highway junction. As a consequence, they arrive at the port an hour late. Traveller A finds out that his ship has left thirty minutes ago, at the scheduled time. Traveller B is told that his ship was delayed by twenty-five minutes and left only five minutes ago. Who of the two travellers will be more upset? In Kahneman and Tversky's study, nearly all participants agreed that this would be traveller B. How did they arrive at this conclusion? A theory-theory explanation would stress the role played by abstract (innate or acquired) intuitions of how different intentional and behavioural states relevant to this scenario are causally related. Given that the desire to catch the ship is equally present in both travellers, the difference must be explained from the dissimilarity between A's belief that, given the heavy traffic, he would have missed his boat anyway, and B's belief that he would still have caught his boat if only the traffic had been just a little less chaotic, or the driver had not missed the junction. Using abstract knowledge of the rules governing folk psychology, one can reason from these represented belief states to the expected degree of "upsetness" and judge whether the answer is A or B. Simulation theorists, in contrast, would suggest that there is no need to apply abstract, folk-psychological rules, or even to represent desires and beliefs, since one can simply compare one's own degree of imagined upsetness in either situation. Defenders of the narrative-practice hypothesis would argue that one's sense of how the projected intentional states in this scenario link together can be correlated with similar scenarios one has acquired previously, leading to one outcome rather than the other. Two-systems thinkers, finally, would leave space for pragmatic combinations of these explanations.

When providing a brief overview of research into mindreading, like I have done in this section, it is unavoidable that each approach is reduced to a simplified sketch giving but an impression of the line of thinking behind it. Clearly, all these approaches have long histories (which also became clear when discussing Premack and Woodruff's 1978 paper) and build on substantial foundations of philosophical inquiry and empirical evidence. Interestingly, it seems that the more papers one reads by (self-)proclaimed defenders of each of the camps, the clearer it becomes that much of the contrasts and controversies are rooted in conceptual and terminological incompatibility, or in dissent with respect to what the relevant questions are, rather than in disagreement over the answers to given questions.[19] One might be reminded of the well-known Indian parable in which six blind men are for the first time confronted with an elephant, and report to one another what this magnificent creature must be like: the one who feels the trunk says an elephant is like a flexible tree branch, the one who feels a leg says it is like a soft pillar, the one who feels the ear says it is like a hairy pancake, and so on. The overarching concept which I have here been referring to as "mindreading" can be found throughout the field under the labels "theory of mind", "folk psychology", "(lower-order) mentalising", "cognitive empathy", "second-order intentionality", and more, all with slight differences in what exactly is meant—not even to mention how this applies to the whole range of adjacent concepts and terms, such as "social cognition", "simulation", "affective empathy", "emotional empathy", "folk-psychological narratives", etcetera. And even abstracting from terminology, there is ample variation in the phenomena and behaviours which are considered to be of interest. In my view, as with the elephant in the parable, real progress will require researchers from different backgrounds to "talk to each other" and cooperate beyond disciplinary borders.

To add just one more example to those already discussed (based on Apperly, 2011: 114-116): imagine a study in which an experimenter sits behind a table that has two boxes on it. A participants sits down at the other end of the

[19] Apperly makes a similar diagnosis, leading to his pragmatic approach of considering a wide variance of existing studies and insights on their merits before laying out his own "two systems"-model. For additional reflection on terminological controversies surrounding "theory of mind" see Schaafsma et al. (2014).

table. The information is provided that there is a piece of chocolate in one of the boxes. Next, the experimenter looks either to the right or the left box, after which the participant is asked to judge whether the experimenter *thinks* there is a piece of chocolate in the box he is looking at. In the first condition, the instructions are such that the participants have to deduce what the experimenter believes and intends in order to locate the chocolate. In the second condition, the instructions are the same, except that one piece of information is added that makes it possible to skip any reasoning about mindstates and simply use the experimenter's gaze as a cue to infer where the piece of chocolate is located. Does this mean that this study is only in the first condition about "mindreading" and in the other condition about, say, "following eye gaze", or "using behavioural cues"? This position is problematic, since in the second condition participants may use either (or both) mindreading and following eye gaze before formulating their answer. Once again this issue can be avoided by using Marr's distinction between the W- and H-levels. On the what-level, it is safe to say that the entire experiment is about mindreading: subjects are asked about where they *think* the experimenter *believes* the chocolate is located. However, on the how-level, generalisations must be made with great caution: it may well be the case that there are differences from one condition to the other, and between subjects, regarding the strategies and mechanisms used to complete the task. This is a point that applies already to Premack and Woodruff's (1978) early paper: after all, they suggest that familiar problems are more likely to be solved through "association", while novel ones require "theorising"—in other words, they already allow room for the possibility that one and the same task involving predictions of others' intentions and behaviours, may or may not require "theory of mind" depending on individual factors and context (see also Apperly, 2011: chapter 6 on this point).

The next two sections provide a further introduction to embedded mindstates or multiple-order intentionality. The focus will no longer be on the H-level of how inferential processes could be carried out when reading minds (as was the case in this section), but will shift back to the conceptual W-level of "what it is" that needs to be explained.

1.2 Embedded mindstates

1.2.1 Multiple-order intentionality

As stated in the Introduction, researchers from various disciplinary backgrounds and convictions have made a case for the importance of the ability to deal with multiple, interrelated intentional states at various levels of complexity. The common way to conceptualise this complexity is using *orders of intentionality*. Originally, the scale of orders of intentionality figured in debates on primate cognition from the 1960s and 1970s. Dennett, who was himself an important contributor to these debates, explained the scale of orders of intentionality as follows:

> A *first-order* intentional system has beliefs and desires (etc.) but no beliefs and desires *about* beliefs and desires [...]
>
> x *believes* that p
>
> x *wants* that q
>
> where "p" and "q" are clauses that themselves contain no intentional idioms. A *second-order* intentional system is more sophisticated; it has beliefs and desires (and no doubt other intentional states) about beliefs and desires (and other intentional states) – both those of others and its own. For instance
>
> x *wants* y to *believe* that x is hungry
>
> x *believes* y *expects* x to jump left
>
> x *fears* that y *will discover* that x has a food cache
>
> A *third-order* intentional system is one that is capable of such states as
>
> x *wants* y to *believe* that x *believes* he is all alone
>
> A fourth-order system might *want* you to *think* it *understood* you to *be requesting* that it leave. (Dennett, 1983 [1962]: 345)

Although this way of counting orders is not as straightforward as it may seem (see Chapter 6), Dennett's explanation does provide a good impression of the

logic of thinking underlying the scale of orders of intentionality.[20] This logic can be detailed in propositions as follows:

P_0	[It is raining outside]	0^{th}-order
P_1	Bill believes that [it is raining outside]	1^{st}-order
P_2	Mary believes that Bill believes that [it is raining outside]	2^{nd}-order
P_3	Peter believes that Mary believes that Bill believes that [it is raining outside]	3^{rd}-order
P_4	John believes that Peter believes that Mary believes that Bill believes that [it is raining outside]	4^{th}-order
P_5	Sally believes that John believes that Peter believes that Mary believes that Bill believes that [it is raining outside]	5^{th}-order
P_n	$Name_n$ believes that P_{n-1}	n^{th}-order

Table 1 – The square brackets indicate that "it is raining outside" is here seen as a fact of the world, independent of a subject having an intentional state about it. In P0 there is no such subject, in P1 there is a subject (Bill) exhibiting first-order intentionality (by having an intentional state about the fact that it is raining), in P2 there is a subject (Mary) exhibiting second-order intentionality (by having an intentional state about Bill having an intentional state about the fact that it is raining), and so on.

The logic of counting orders of intentionality in this way has inspired a vast amount of research in experimental psychology and cognitive neuroscience, ranging from the development of tests to assess individuals' performance on reasoning tasks involving varying orders of intentionality, to a focus on typical and atypical development, involved brain areas, and formal models of the

[20] Incidentally, below I will discuss a problematic side of this logic that seems to some extent prompted by the very term *"order(s)"* of intentionality. The idea of it being "orders" evokes questions such as "How many orders can a species/individual process?" or "What is the number of orders involved in this task/event/story?" I will argue that intentional states are most of the time not "piled up" (as orders), but interlinked in various kinds of ways. In this Section and in other parts dealing with or building on the long tradition of research on this topic, I will retain the term multiple-order intentionality. In Chapter 3 I will discuss my alternative concept of the "thoughtscape" in more detail.

mechanisms underlying the ability to deal with embedded mindstates.[21] In addition, linguists, literary theorists, archaeologists, anthropologists, and researchers from a handful of other fields have used the concept of embedded orders of intentionality in their frameworks (various examples will be discussed in Sections 1.2.2 and 1.3 below). In this thesis, the totality of research that has implemented the logic of the orders of intentionality in some form will generally be referred to as research within the *mentalising paradigm*, named after the tests used for the assessment of one's competence to reason with embedded intentional states, the so-called "mentalising tests".

In mentalising tests, subjects are asked to read or listen to short stories describing a particular sequence of social interactions, such as the organisation of a surprise party. The story is followed by questions of the form "Did A know that B wanted C to come to his party?", or "Did C know about the party?", or "Did B want A to think that C should know about the party?". By using three to five such stories, each followed by around ten questions of differing orders of complexity, a score indicating "mentalising capability" can be calculated for each individual participant. In a range of studies, scores from this test have been shown to be associated with various sorts of measures of people's social capabilities and real-life social functioning. For example, a number of studies have indicated that mentalising scores correlate with estimates of social network size, suggesting that those participants who perform better at mentalising tests have, on average, more people in their social networks. Another study has indicated that participants with higher mentalising scores were less likely to attribute causes of negative events to others: they appeared to be, as it were, less "distrustful" of others' intentions in a social context. Other studies have investigated the relations between mentalising, empathy, and executive functioning, or mentalising skills and language competence. Also, a version of the mentalising test adapted for children showed an association between test scores and general social aptitude as assessed by their teachers. Another perspective was added by various studies in the field of social

[21] For a discussion of research into typical and atypical development see Baron-Cohen et al. (2013); for involved brain areas see, among others, Frith and Frith (2003) and Rushworth, Mars, and Sallet (2013) and Mars et al. (2013); for formal models see, for example, Behrens et al. (2009) and Yoshida et al. (2008; 2010). See also Note 14 above.

neuroscience: higher mentalising scores have been shown to correlate with higher amounts of grey matter in cortical areas important for social functioning.[22]

All these statistical associations may be taken to indicate that mentalising tests *do* tap into at least some skills and properties relevant to actual social life and interaction. However, authors presenting mentalising research themselves, as well as critical outsiders, have stressed that it is still to a large extent unclear *why* these associations exist, or in other words: little is known about which mechanisms are targeted by these tests and how precisely they relate to real-life social interaction. In addition, discussions have arisen over ecological relevance and methodological soundness of the questionnaires, but the tests have been improved over the years and researchers have found ways to control for factors such as general memory capacity or language ability.[23] Throughout this thesis, questions pertaining to the mentalising tests will return in various forms, and a detailed analysis will be offered in Chapter 6.

1.2.2 The roles of language

Earlier in this chapter, the option of formulating explicit mindreads has been discussed (Section 1.1.1), along with the possibility to describe and categorise different mindstates and their mutual relationships using linguistic propositions (Sections 1.1.2 and 1.1.3). With this, however, only one of three roles of language in relation to mindreading has been addressed. The current section

[22] The "mentalising test" (also sometimes referred to as the "Imposing Memory Task" or "IMT") was originally designed by Kindermann, Dunbar, and Bentall (1998), for a study in which they investigated the relation between test scores and causal attribution of negative events. Afterwards, the test was revised, updated, and adapted several times. Stiller and Dunbar (2007) demonstrated a positive correlation of mentalising scores with estimates of social network size, which was replicated several times (see Lewis et al., 2011; Powell et al., 2014; Launay et al., 2015). All these studies suggest a better performance among women. For research showing associations between mentalising performance and volume of the orbital prefrontal cortex see Powell et al. (2010), Lewis et al. (2011), and Powell et al. (2014). See Launay et al. (2015) for mentalising in relation to empathy and executive functioning. For mentalising in children see Liddle and Nettle (2006) and in adolescents see Haddad (under review). An elaborate analysis of these studies is offered in Chapter 6 of this thesis.

[23] See Launay et al. (2015) for a general discussion, O'Grady et al. (2015) for a critical review of the methodology and an alternative testing method using movie clips, and Oesch (2015) for mentalising and language competence.

distinguishes these roles and points forward to the chapters in which they will be discussed in more detail.

The first role of language is thus the *representation* of mindstates and mindreading tasks. This can itself be subdivided into *formal* or *propositional representation*, where mindstates and their mutual relationships are made explicit for the purposes of investigating them and assessing their complexity (as in, for example, Table 1), and *natural representation*, the way in which mindstates and their relationships are rendered and managed in various genres of natural discourse, including novels, plays, newspaper texts, radio reports, conversations, etcetera. Language in the role of representing mindreading, both propositionally and naturally, is important throughout this entire thesis. In Chapter 2 and 3, the focus will be on (literary) narrative language, in Chapter 4 and 5 more everyday forms of language usage (newspapers, conversations) will enter the stage, and Chapter 6 will deal with linguistic representations of mindreading tasks in the context of psychological experiments.

The second role concerns the *conceptual support*, *scaffolding*, and/or *training* that language can provide for our mindreading skills, even when the actual reasoning is performed implicitly and/or non-linguistically. For example, various researchers have suggested that children around the age of 3-4, who learn to deal with embedded sentences (e.g. "Snoopy thinks that the candy is in the box"), not only acquire a way to *communicate* about mindstates and perspectives, but also learn a formula for *thinking* about them in the first place (see e.g. Lohmann and Tomasello, 2003; Milligan et al., 2007). In other words, the matrix structure of such sentences may not only provide a new "label" for an existing reasoning process, but also add a new strand of reasoning to a child's thought repertoire. In a similar way, stories can be argued to form a natural training environment for one's mindreading skills. They offer insight in the fictional minds of characters, thereby enabling one to experience "what it is like" to be inside someone else's head, and they provide a mode for projecting hypothetical social scenarios, thereby avoiding the potential costs of trying these out in real life. On top of these ways of support, scaffolding, and training

for mindreading, as identified by various authors,[24] I suggest an additional one, which pertains to the *structural* properties of narrative language usage. In short, as I will argue in Chapter 2, narrative language features all kinds of strategies for fleshing out perspectives and mental states, and for mutually coordinating them in a natural and comprehensible way. Learning to deal with narrative may therefore hone one's "real-world" capabilities of switching between multiple perspectives, understanding situations in terms of the underlying perceptions, intentions, motives, etcetera, and mapping behavioural patterns on particular mental states. Viewed this way, narrative is not just a way of speaking, but also a way of *thinking*, which is at least partly governed by the conventions of narrative language that we acquire in the context of learning to understand and tell stories (see Van Duijn, 2015). More details on this idea will be worked out in Chapter 2, 4, and 5.[25]

Not only does language thus serve to represent mindstates and mindreading tasks (first role), nor is it just likely to provide implicit support and scaffolding for our mindreading abilities (second role), it is also in important ways *itself dependent on* and *building upon* mindreading. This third role of language makes things complicated: after all, if all three are considered together, it is implied that language and mindreading must have a relationship of mutual dependency and "cosupport" in developmental terms, and one of "coevolution" in evolutionary terms—which is precisely what I will assume throughout this thesis, and argue for in various ways. Such arguing is necessarily incomplete and to some degree speculative, given the issue's enormous psychological complexity and evolutionary depth of hundreds of thousands of years. Nonetheless, I hope to provide convincing arguments and evidence at various points that it is the *only* possible way of construing the relationship between language and mindreading. With an eye on that, it is important to briefly introduce some concepts from the study of human interaction, with which I will round off this section.

[24] For a variety of views in the broader area of literature and (social) cognition see, among others, Zunshine (2006); Boyd (2009); Vermeule (2011); Oatley (2011); Nussbaum (2011); Djikic et al. (2013); Carney et al. (2014).
[25] Note that this idea is reminiscent of the "narrative practice hypothesis" (e.g. Gallagher and Hutto, 2008), but only partly overlaps with it; see also Section 1.1.4 above.

Human interaction is, broadly, the context in which language usage takes place. In the default version it happens face-to-face between a S(ignaller) and A(ddressee) who reverse roles with every turn taken, using a multimodal stream of auditory, visual, and palpable cues—all other interaction forms, such as writing, phone calls, text messaging, and so on, are ultimately variants of or derivatives from this default setting (Fillmore, 1981). Interaction is by no means always linguistic: humans can manage each others' behaviour, share information, make friends, play tricks and jokes, and interact in all kinds of other ways without ever saying or writing a word. This is known to anyone who has ever been "lost in translation", trying to get around in a place where no one speaks one's language. Or, another good example of how rich interaction can be without the aid of language is provided by the game of charades: players often manage strikingly well in getting complex meanings across, even though all conventional, mostly linguistic symbols are banned (except for a few ones specific to the game). However, any player of charades or anyone being lost in translation also realises how strained and impoverished communication without language is. To use an amended version of Scott-Phillips' (2015: 16) words: language is not what makes interaction possible, but what makes it powerful.[26] After all, as will be discussed in more detail in Chapter 5, the conventions of a language can be seen as "supercues", coagulated local solutions (i.e. within one cultural-linguistic community) to the coordination problems that arise when interacting. In this view, every lexical item and grammatical procedure ultimately is the result of generations of language users trying to coordinate their mindstates in interaction with each other and the environment, thereby converging on solutions that are communicatively effective, physically and cognitively efficient, and learnable for new generations of language users (see also Verhagen, 2015; Mesoudi, 2011; Tomasello, 2008: chapter 6).

In his analysis of the distinctive properties of human interaction, Levinson (2006) introduces the concept "Schelling mirror world". Schelling was an economist who studied a specific species of coordination problems: the ability of subjects to arrive at a solution together in the absence of

[26] The full version of Scott-Phillips' quote will be discussed in Chapter 5 (and contested on an important part not cited here).

communication. For example, if they are told that they have to meet someone else in Moscow the next day, but not exactly where and when, and they know that the other has had precisely the same instruction, they can perform much better than chance would permit by (implicitly or explicitly) asking themselves what the other will think, and what the other will think that they will think. A "Schelling point" (Schelling, 1960) high above the odds in Moscow is probably "12 noon at the Red Square, in front of the clock tower besides the Kremlin". If one has to meet in a theme park, this point would probably be the entrance, or in a crowded department store it may be the "lost-and-found" desk. Converging on such Schelling points, according to Levinson, requires not only a special way of reflexive thinking (about what the other will think one will think, etcetera), but also a notion of mutual knowledge or *common ground*, including a sense of mutual salience: "what leaps out of the common ground as a solution likely to independently catch our joint attention" (2006: 49, referring also to Clark et al., 1983, and Clark, 1996). He argues that these same ingredients are also requirements for human communication: reflexive thinking and common ground, including a mutual sense of salience. After all, as has been described by many linguists and philosophers of language, there are thousands of possible ways in which a particular meaning can be expressed, while at the same time, every expression can have many different meanings.[27] Only through the same combination of reflexive thinking and common ground, including a sense of mutual salience, can humans coordinate their mindstates while interacting, or in Levinson's words: it is through these factors that "meetings of the mind" can occur in the "Schelling mirror world" that underlies human interaction (2006: 49; for an experimental approach see Stolk, Verhagen, and Toni, 2016; Stolk, 2014).

Grice (1957) was the first to present a fundamental study of how communicative meanings can arise despite the indeterminacy of linguistic

[27] This can be demonstrated using nearly any utterance, but consider the example of me saying to a friend: "hey, there is Ann". If we are standing outside a music venue, and Ann has our tickets, this probably means something to the effect of "all right, we can go inside". However, if Ann is my friend's ex-girlfriend and we are about to enter a bar for a drink, it can mean "let's go somewhere else"—unless my friend has just told me that he hasn't seen his ex-girlfriend in a while and would be interested in a conversation with her, in which case it probably means "what a coincidence, let's go inside"... etcetera. For a discussion see, among many others, Keller (1995), Sperber and Wilson (1995), Clark (1996), and Scott-Phillips (2015).

expressions as such. According to his theory of meaning, "a signaller S communicates z by behaviour B if S intends to cause an [addressee A] to think z, just by getting [A] to recognise that intention" (Levinson, 2006: 49; "recipient" in original replaced by "addressee"). Sperber (1994; 2000) and Scott-Phillips (2015) have reformulated this insight in terms of a multiple-order mindreading problem, suggesting that for any full-blown linguistic interaction event:

S intends
that A should recognise
that S intends
that A should believe
that z

The precise nature of this mindreading problem, assumed to be at the heart of language usage, will be detailed (and contested) in Chapter 5. The version at which I will eventually arrive, building on Clark's (1996) and Verhagen's (2015) notions of *common ground* and *joint intentionality*, suits the lazy mindreader by being much more economical in terms of the assumed amount of cognitive complexity. In short, it turns the argument upside down: instead of suggesting (following Sperber and Scott-Phillips) that interaction works because interlocutors (somehow, implicitly) take the steps spelled out above in order to "meet" each other at five orders of embedded intentionality, I argue that *as a rule* they start off having already met—and instead of suggesting that it is necessary by default, I suggest that it is only in exceptional cases that such steps need to be taken (for example, as will be discussed in Chapter 5, in order to work out and repair a misunderstanding: "Ah! I thought you intended me to think that….etc.").

Put differently: *in theory* it is possible for interlocutors to reflect on the communicative situation in the way suggested by Sperber and Scott-Phillips, but in practice it is rarely necessary. Normally, a signaller "tosses" a particular behaviour (typically a string of sounds, gestures, and facial expressions) into the Schelling mirror world, assuming that the addressee will be able to figure out what the signaller means by it. In nearly all instances of communicative interaction there are several principles and mechanisms at work that save the

signaller and addressee from having to apply multiple-order mindreading. Summarised in brief:

- Common ground/joint intentionality: interlocutors always start from a set of shared beliefs or "common ground" (Clark, 1996) instead of having "join" their individual sets of intentional states each time they interact;
- Ready-mades/packages: for many expressions, occurring in particular contexts, we may have existing meaning associations stored in our memory that are either shared between speaker and addressee in particular, or among members of the cultural-linguistic community more widely. Such associations can be easily retrieved, compared, adjusted, and used as ready-made blueprints or frames in interaction, without having to establish complex meanings "from zero" (as worked out for examples such as "allegedly" and "accidentally" in Chapter 4).
- Interactive structure/alignment: in interaction we do not have to sort everything out by default and right away—in every communicative turn we seem to build representations that are "good enough" for the interaction to keep going, but no better (cf. Apperly, 2011: 114-119). If required, interlocutors can work out a particular point in more detail, aiding and steering each other in the desired direction turn by turn. Many conversations do not have "signal-response" as their basic structure, but rather "testing-adjusting-retesting" (Levinson, 2006).
- Relevance: driven by the need for communicative efficiency, signaller and addressee have both learned from their experience as communicators to become geared towards choosing maximally relevant solutions. This means that, in most cases, what the signaller has to do is pick the first expression that comes to mind, while the addressee has to pick the first interpretation that comes to mind. If this does not work, they can try the second-most relevant expression or interpretation; thus both speaker and addressee in practice work downwards on the gradient of relevance (cf. Apperly, 2011: 115-116, referring to Sperber and Wilson, 2002; and Chapter 5).
- Ratchet effect of linguistic items: not only are signaller and addressee experienced in choosing the most relevant cues and interpretations, the

> linguistic tools they have available also store a wealth of such accumulated "experience". After all, they have emerged as a result of numerous instances where generations of signallers have tried to get particular meanings across to addressees, in settings that have for at least some important parts not changed (cf. Chapter 4 and 5).

As I will argue at several places throughout this thesis, in many cases of daily interaction these mechanisms work so well, that mindreading in the "full" form as suggested above is hardly ever needed in order to communicate—it is only in exceptional cases, such as when trying to repair a misunderstanding, playing a sophisticated pun, or reflecting on the very act of communication, that participants in a communicative setting are incited to go "all the way down" and work out what the other intends that they understand that the other wants...etcetera. In other words: language is what makes human interaction so powerful not just because it can represent mindstates and their relations in efficient ways (first role), nor just because it may support mindreading implicitly (second role), but also because it can work as a "mindreading-avoidance tool". Mindreading is indeed necessary for communicative interaction (third role), however, various mechanisms and principles that are part of, mediated by, or closely tied to language save interlocutors the trouble of having to process all steps suggested by Grice, Sperber, and Scott-Phillips by default.

1.3 The social brain

1.3.1 Early primate roots

Up to now this chapter has been concerned with the fundaments of what intentional states are like, how mindreading can work, and how both relate to language. However, as stated in the first section, there is another basic question: why is it that we bother about mindstates of others at all? The context in which an answer to this question can be provided (and, indeed, the context in which this question itself becomes relevant) is offered by research surrounding the

social brain hypothesis. At the core of this hypothesis lies the idea that the complex social environments in which primates have lived in their evolutionary past were the primary drivers behind the emergence of their increasingly large and powerful brains—or rather, *our* large and powerful brains, since humans are of course included in the primate order.[28]

The briefest version of the story of primate evolution goes as follows. In the geological period known as the Palaeogene or Lower Tertiary, a bit over 50 million years ago, certain mammals on the African continent started foraging in groups (see Shultz, Opie, and Atkinson, 2011, for key evidence supporting this scenario). This development was probably driven by a transition from nocturnal to diurnal activity, which increased the risk of being attacked by predators while moving around in search of food. Although living in groups lowered vulnerability to predation, at the same time it posed some very particular challenges for the ancestral primates, including finding new ways of organising reproduction and care for offspring, resolution of conflicts arising over access to resources within groups, avoidance of the costs inflicted by freeriders, and coordination involved in moving collectively or protecting the group against risks from outside. In response to these challenges, various species have evolved different solutions over millions of years of time, as reflected in the different forms of social organisation and complexity that can be found throughout the primate world today (for references and more detailed overviews of the early episodes in primate history see Dunbar, 2014: chapter 2 and 4; Gamble, Gowlett, and Dunbar, 2014: chapter 3 and 4).

An important characteristic of social organisation found throughout the entire primate world is the tendency to form intense social bonds and coalitions

[28] The suggestion of the association between brain size and social complexity was originally made by Jolly (1966) and Humphrey (1976), before it was addressed in Byrne's & Whiten's volume *The Machiavellian Intelligence Hypothesis* from 1978. Dunbar further developed the social brain hypothesis and was the first to test it systematically (1992; 1998), discovering the correlation between social group size typically formed by a primate species and its neocortex size, and subsequently presenting numerous findings supporting and/or refining the hypothesis. He has written a vast amount of publications on the topic, several of which play important roles throughout this thesis. For this introductory section, I will to a large extent follow the line of his recent overview book *Human Evolution* (2014), along with the co-authored volume *Thinking Big. How the Evolution of Social Life Shaped the Human Mind* (Gamble, Gowlett, and Dunbar, 2014), only at key points referring to the original papers. In the rest of this thesis, the original papers will be used.

that last over longer periods of time. These bonds and coalitions directly or indirectly protect individuals against the challenges and costs of living in groups and defuse the stress that comes with it. Primate bonds have an emotional and a more cognitive component. The first is primarily mediated through endorphins triggered by specific social activities. The latter can be defined in terms of having a mutual sense of trust and obligation, and, in some cases, willingness to provide help and support, all of which require some form of cognitive coordination (cf. Dunbar, 2014: chapter 2). I will return to the cognitive demands of social living below, but first focus briefly on the emotional, endorphin-mediated component.

1.3.2 The "bonding gap"

The main activity associated with endorphin release in non-human primates is social grooming, the process where one individual sifts through another's fur to remove small bits of debris and inert skin. Besides hygiene benefits, this mildly painful treat triggers the release of endorphins in the brain, which alleviate stress and pain levels and presumably underpin the feelings of emotional closeness that we know from friendship and love. This mechanism is still at work in humans, as can be sensed when receiving a massage or engaging in light stroking and cuddling. However, in our case, time has seen the addition of other mechanisms of maintaining intense social relationships, which have taken over much of the heavy lifting (Dunbar, 2014: chapter 1 and 8).

According to Dunbar the transition from social grooming to other bonding activities constitutes one of the main threads in the evolutionary story of our lineage. Simply put, when group sizes increased, our ancestors must have run up against time limits: since grooming is an inherently time-consuming, one-on-one activity, it works fine for smaller groups, but will put high pressures on time budgets in larger groups. A partial solution found in some species of primates is to invest in a strong relationship with a few core social partners (instead of weaker ties with many or all group members), thus breaking up the larger group into interlinked and partly overlapping coalitions. This structure was most likely found in the groups formed by our hominid ancestors, and

arguably is still visible in present-day human social life.[29] However, it was estimated that if ancestral human societies had relied on strictly one-on-one bonding activities such as grooming, their members would have needed to spend over forty per cent of their day doing this, which would have conflicted severely with the time budgets reserved for foraging and resting (Dunbar, 2014: chapter 7; Gamble, Gowlett, and Dunbar, 2014: chapter 5). In other words, it seems likely that the amount of free time left for social activities *after* foraging and resting, put a constraint on the maximum number of social ties individuals could maintain. Therefore, in order to be able to break through various glass ceilings of maximum group sizes, more time-efficient bonding activities had to emerge in our lineage, bridging the "bonding gap" (Dunbar, 2008) between the groups of 40-60 individuals, in which our early ancestors lived, and the groups of around 150, as formed by anatomically modern humans—and this is precisely what Dunbar and colleagues have argued: activities involving for instance laughter, dance, music, and, of course, language have become our alternatives for social grooming.

Both dance and laughter have the capacity of fairly straightforwardly triggering endorphin release, thus supporting emotional social bonding in a direct way. A similar case can be made for singing together. Also, importantly, dancing, singing, and laughing can be done together with several others at the same time, greatly increasing the effectiveness of time spent socially.[30] The same, of course, holds true for talking. However, the links between language and social bonding are more complex. Talking and listening *as such* do not seem to be contributing much to social bonding: rather, language contributes indirectly through such activities as gossiping and sharing jokes, myths,

[29] For this social group structure, referred to as the "fission/fusion-model" see Dunbar (2003).
[30] For an overview see Dunbar (2014: chapter 2, 3, and 8). For social bonding in relation to dance and moving "in synchrony" more widely see Tarr, Launay, and Dunbar (2014); for laughter and social bonding see Dunbar et al. (2011); for singing see Pearce et al. (forthcoming). Research shows that laughing, even today, is typically done in intimate cliques of two to four individuals, rather than in larger groupings. Interestingly, these cliques are similar in size to the groups people tend to form in natural conversations (see Dunbar, 2015).

religious stories, and fiction.[31] However, with the possible exception of telling (some sorts of) jokes, these activities all require a highly sophisticated form of language to be in place, capable of representing at least some abstract concepts, referring to events outside the here-and-now, and coordinating multiple referents and possibly their mindstates (cf. also Tomasello, 2008: chapter 6). The emergence of such sophisticated language forms is typically assumed to be of a relatively recent date in our evolutionary history (possibly only with the arrival of *Homo sapiens* around 200,000 years ago or even later; see e.g. Fitch 2010; Perreault and Mathew, 2012), whereas our social group sizes, and thus our need for efficient social bonding mechanisms, have shown important increases much earlier, going back probably around 2 million years (Gamble, Gowlett, and Dunbar, 2014: chapter 5). This would suggest that other factors (such as dance, music, and laughter) were at first more important in bridging the bonding gap, while various forms of gossiping and storytelling came in later. In addition and related to this, as discussed in Section 1.2.2 above, it is argued by some researchers that mastering a language "as we know it" requires powerful mindreading capacities, which is also considered to be a reason for why the emergence of such sophisticated language must be dated to our more recent history.

The first of two main ways in which social group size can be linked to cognition, and to mindreading in particular, can thus be summarised as follows: bonding larger groups may require language and storytelling skills, which rely on mindreading capacities, which again rely on large and powerful brains. However, one of the objectives of this thesis is to rethink the way in which mindreading and language are related, so this topic will be continued (and the argument partly challenged) at various places, especially in Chapter 5 and in the Conclusion.

[31] Besides contributing to endorphin-mediated bonding indirectly, language clearly has other important advantages in the context of keeping increasingly large communities together: talking can "time share" on other activities (such as walking, eating, and cooking) and it can be used to share information about the social network (gossip) in a much more efficient way than by personal observation (see Dunbar, 2014: 227).

1.3.3 Cognition and primate social life

Besides the emergence of new bonding activities, a second main thread in the story of human evolution identified by Dunbar is brain size. The social brain hypothesis projects that if an animal's social life is more complex (and thus demands more sophisticated social behaviour), it will have more grey matter in the brain areas associated with social cognition. Support for this relationship has been found throughout the entire animal kingdom, both at the level of one species compared to another species and at the level of individuals within the same species. Dunbar (2014) discusses research showing that it holds even in social insects: species with more complicated social structures show increased "brain" volume (or relevant neural network size), compared to species of social insects with less complex social structures. In a similar vein, queen bees have a significantly more sophisticated social life than their worker sisters, and also show more relevant brain volume. Species of birds forming pairbonds, hence needing to be able to maintain intense, long-term relationships with their partners, have larger brains compared to birds who have more flexible mating systems. Primates with more diverse repertoires of social behaviour, for example involving deception or alarm calls, tend to have relatively larger neocortices. The same is true for primates living in larger social groups, where they have to maintain higher numbers of social relationships and/or exhibit more diverse repertoires of social behaviour. This is reflected in the correlation between mean group size and neocortex size as plotted in Figure 3:[32]

[32] Note that the social brain hypothesis is thus essentially about social complexity, and not about the number of relationships an individual can deal with per se. "Qualitative" factors are equally important: for example, mating strategies, deception rates, or coalition complexity all correlate with relative brain size irrespective of total group size (Shultz and Dunbar, 2014: 49-50; cf. also Dunbar, 2008).

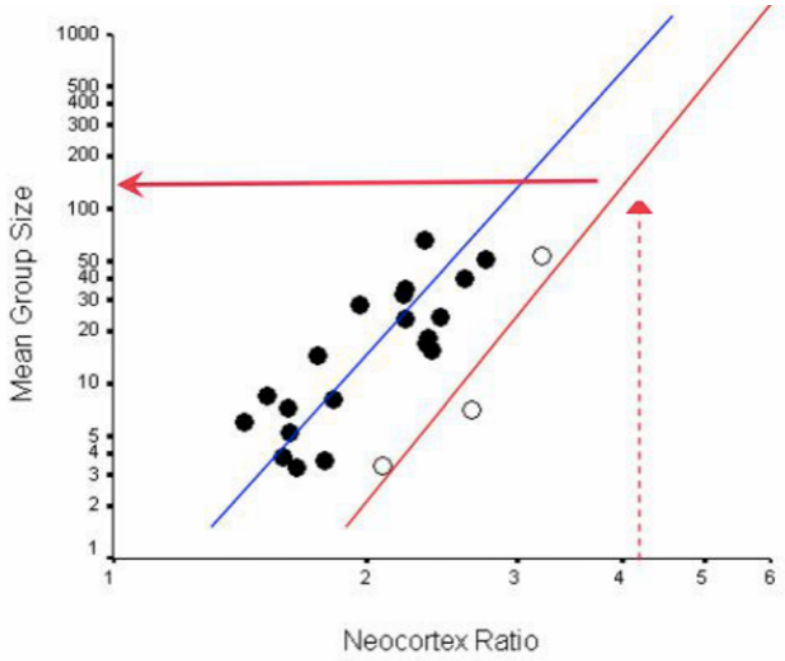


Figure 3 – The filled dots represent species of prosimians and monkeys, showing a robust correlation between mean group size and neocortex ratio, the size of the neocortex divided by the total brain size. The open dots represent the three species of great apes, from left to right orangutans, gorillas, and chimpanzees. The red arrows indicate the predicted mean group size for modern humans, based on their neocortex ratio factored into the ape equation: ±150 individuals, known as "Dunbar's number" (Dunbar, 2003).

Group sizes in our primate ancestors were not only constrained by the availability of social time for grooming and/or alternative bonding activities (which may or may not have involved mindreading, as discussed in the previous section), but also by cognitive limits in a more direct way. This is the second of the two main ways in which cognition, and mindreading in particular, can be linked to social group size. In order for a group not to break down into chaos, individuals need to *coordinate* their behaviour with respect to one another: from the very basic capacity to move in accordance with others' movements, a skill apparent in for example bird flocks or ungulate herds, to sophisticated empathic, cooperative and strategic behaviours, as in for example targeted helping, conflict resolution, group hunting, deception, or consolation. Mindreading competence at various levels of complexity has been argued to

underpin different aspects of such coordination of individuals within groups, each of which will be discussed in the subsections below.

1.3.4 Mindreading, coordination, and group size

In the case of birds flying in a flock, coordination of behaviour from an individual's perspective comes down to adjusting to a few others that happen to fly near. The apparently sophisticated movements emerging on the level of the flock thus do not require birds to bother about anything beyond the movements of a handful random neighbours: there is no need to distinguish between them individually, and relationships with them do not have to persist beyond the coordination task itself (David-Barrett, 2014).

Figure 4 – Individual birds in a flock only coordinate with a handful of neighbours, but do not need to have a representation of the behaviour of the flock as a whole (David-Barrett, 2014).

As a consequence, little cognitive investment has to be made for successful group membership and the cognitive costs do not add up when group size increases. This stands in stark contrast to the situation in primates, where individuals in principle maintain a constant relationship with every other individual in their social group. When interactions of some form occur, this existing relationship functions as the basis, instead of the interaction being the

basis for an ad hoc relationship (as in flocks and herds). Therefore, primates have to be (and, indeed, *are*, Parr and De Waal, 1999) able to distinguish between individual group members and, to a certain extent, they have to keep track of previous interactions they have had with them. This is an important reason why primate group life causes cognitive load: when groups grow larger, their members have to tell apart more individuals and remember the current status of their relationship with each of them.

Moreover, there is an even more weighty reason why primate sociality is cognitively demanding. In order to fit into a primate group, it is not sufficient to know one's own relationship to all other group members: one has to keep track of "third-party relationships" between all of the other group members as well. For this reason, the number of relationships primates have to be able to distinguish and keep track of, in theory, can be shown to increase exponentially with every extra group member added (David-Barrett and Dunbar, 2013). Evidence from observations of social structures in many different primate species, in relation to their relative brain sizes, supports this idea (Shultz and Dunbar, 2014).

However, all of this does not yet warrant the importance of the capability to deal with some form of mindreading: a primate individual can in principle monitor social relationships within its group without having to deal with intentional states. Mindreading seems to come in as soon as *prediction* of others' behaviour and intentions enters the stage—yet in this case it is a possibility, not a necessity. For example, when making decisions about whether or not to act cooperatively towards another individual, one will have to predict whether the other will reciprocate this approach, or exploit it. The scenario of deciding whether to cooperate or not has therefore been linked to the ability to deal with multiple orders of intentionality. Yoshida et al. claim that such a decision is necessarily based on "recursive representations of another's intentions, since if I decide to [cooperate], I must believe that you believe that I will cooperate with you" (2010: 10744). There seems to be ground here for an arms race inflicting a constant pressure on individuals to stay ahead of their peers. Advantage can be gained if one can from time to time make the other believe that one intends to cooperate, while in fact one is about to exploit him. At the same time, it is important to be able to detect when the other intends one to believe that he will

cooperate while in fact he will not… (Byrne and Whiten, 1988, have emphasised the importance of such "Machiavellian" cheating and cheater detection in the evolution of primate social intelligence; see also Tomasello, 2014: 37-38).[33]

Note that I wrote purposefully that it *can* be the case that mindreading is needed when prediction of others' behaviour and intentions enters the stage, but that it is not a necessity. A point that often seems to be overlooked in studies that link mindreading to cooperation is that calculating whether the other believes that one intends to cooperate is one way of predicting the other's reliability as a cooperation partner, but surely not the only way. Another option would be to make this prediction on the basis of past behaviour of the other, either towards oneself or towards others. All that is needed is the ability to tell individuals apart and a memory for previous interactions, but no mindreading (see Wilkinson, 1984, for an example in vampire bats). A third way of predicting another's intention to cooperate would be using emotional or situational cues: is he nervous, are there more group members around to watch his behaviour, etcetera. A fourth way would be to make an assessment of the "rationality" of the task: what can the other gain from cooperating, and what does he have to invest? In fact, it would make most sense to use combinations of these ways (and potentially even additional strategies), and it may well be that this is what individuals do in practice. In short, deciding whether or not to cooperate with others may involve mindreading, but not necessarily so.[34]

---

[33] Another way in which cooperation and mindreading can be linked is by factoring in third-party opinions: the question for X whether or not to cooperate with Y can also depend on predictions of what Z would think of this alliance. This adds complexity in terms of the number of intentional states involved in making a decision for X, without these intentional states necessarily being *embedded* (Y thinks… and Z thinks…). The issue of interlinked (but not embedded) intentional states will also return in various forms throughout the following chapters.

[34] In some way this comes down to saying, more generally, that not all mindreading tasks are solved through mindreading; or more precisely: some tasks generally considered to involve mindreading (such as decisions about cooperation) can be solved *both* in ways generally considered to be mindreading (e.g. placing oneself in the other's shoes, reasoning about the other's possible motivations) and in ways generally not considered to be mindreading (e.g. extrapolating from the other's previous behaviour). Note that this is the same point as made at the end of Section 1.1.4., and that this again highlights the importance of distinguishing between different levels of explanation. Sometimes it can be said on the W(hat)-level that "X decided to cooperate because he *thought* that Y *intended* to cooperate as well", whereas on the H(ow)-level this decision came down to (for example) mere extrapolation of X's experience that Y always cooperated on previous occasions. Is this a case of mindreading? Yes on the W-level, no on the H-level. (Cf. also Kümmerli et al. 2010).

Apart from the issue of deciding whether or not to cooperate, does cooperation itself involve mindreading? For example, chimpanzees seem to hunt cooperatively (e.g. Boesch, 2005), which could be argued to require A to *understand* that B *wants* to move around the tree, so that B *intends* A to *understand* he should take the other side, and so on, implying mindreading at multiple levels of complexity. However, it has been pointed out that a more likely scenario is that all chimpanzees participating in a hunt try to maximise their own chances of catching the prey, which results in a situation that only *seems* to be coordinated intentionally from the perspective of an outside observer (similar hunting is found in hyenas, lions, and wolves; see Tomasello, 2008: 173-175; Dunbar, 2014: 244). Such hunting "alone together" does not seem to require much mindreading, apart from again the possible Machavellian twists of misleading others or anticipating potential misleading by others. In that sense, this form of cooperation may on the level of its underlying mechanisms well come closer to the bird flock than it seems at face value.

1.3.5 Mindreading and social learning

It has further been claimed that mindreading underpins living in social groups through facilitating effective learning mechanisms. Whereas some theorists have suggested that most of the important learning takes place through imitation, and therefore does not rely on taking others' perspectives (see Heyes, 1993; 2012a), Tomasello and others have suggested that it is precisely because we, humans, are able to picture ourselves in someone else's shoes, that we can learn "through" them (Tomasello 1999; 2008; 2014). In this way, thanks to our mindreading competences, cultural conventions can reliably spread through a group at a fast pace, since learning not only takes place from parents and caretakers to a new generation ("vertically"), but also "horizontally" between peers:

> The form of social learning required here is not just imitation, but role reversal imitation, in which each initiate to the convention understands that she can use the convention toward others as they have used it toward her, and vice versa—so that both producer and comprehender

> roles are implicitly present in both production and comprehension.
> (Tomasello, 2008: 221-222, referring to Tomasello, 1999)

Groups of humans (and potentially some of our hominid ancestors and ape relatives)[35] having such a mechanism in place through which conventions spread, turn into cultural communities where coordination between group members works in a highly effective way: when conventions are mutually shared, there is (as it were) a supra-individual order capable of *orchestrating* behaviours and interactions in all kinds of domains. This saves huge amounts of negotiation and trial-and-error costs—time, risk, energy, cognitive power, and so on—otherwise borne by individual group members. This will be elaborated further in Chapter 5, where the notions of "joint intentionality" and "common ground" are introduced (following Clark, 1996, and Verhagen, 2015) and where linguistic items will be viewed as coagulated solutions to coordination problems occurring when interlocutors try to update a set of shared beliefs.

1.3.6 Mindreading, language, and narrative

Apart from (but clearly related to) cultural learning, the ability to deal with multiple orders of intentionality has been argued to enable and support language ("third role" in terms of Section 1.2.2 above), and thereby activities important for living in social groups, such as gossiping and storytelling, as discussed above. Dunbar (2014) and others consider the latter activities highly important factors in how our hominid ancestors could break through glass ceilings of group size and brain capacity. The final section of this chapter will be concerned with Dunbar's view on the role of language and stories in the context of the social brain hypothesis. This brings the discussion back to the

[35] There is evidence that chimpanzee groups also have some form of cultural conventions that spread both horizontally and vertically. However, compared to the human situation, there clearly is an enormous difference in the amount to which these conventions modify and enhance the chimpanzee ways of living. See Whiten et al. (1999) and Whiten (2011).

core issue of this thesis: the relation between mindreading, language, and narrative.[36]

As stated earlier, the basic idea advocated by Dunbar is that our lineage, over time, exhibited increasingly better mindreading competences. In brief, it is assumed that our current capacity comprises five "levels" of intentionality (Kinderman et al, 1998; Stiller and Dunbar, 2007) and that the last ancestor we shared with our closest relatives in nature, chimpanzees and bonobos, could, like them, achieve at most two of such levels. Smaller-brained monkeys are assumed to be capable of only one level. Combined with the mentioned neuroimaging experiments suggesting that, in human subjects, mindreading competence is correlated with brain mass in areas relevant to social cognition, a function can be hypothesised expressing brain size in terms of achievable level of intentionality (Powell et al., 2011; Dunbar, 2014: chapter 7). When brain sizes of our ancestral hominids, estimated on the basis of fossil skull bones, are factored into this function, this yields the following graph:

[36] This thesis will not explicitly address religion, but it is clear that religious traditions rely for an important part on the exchange of stories. Therefore, much of what will be said in this thesis about stories in relation to mindreading is also relevant to building and maintaining religious communities. For a discussion of religion in relation the orders of intentionality, see Gamble (2010) and Dunbar (2008; 2014: chapter 8).

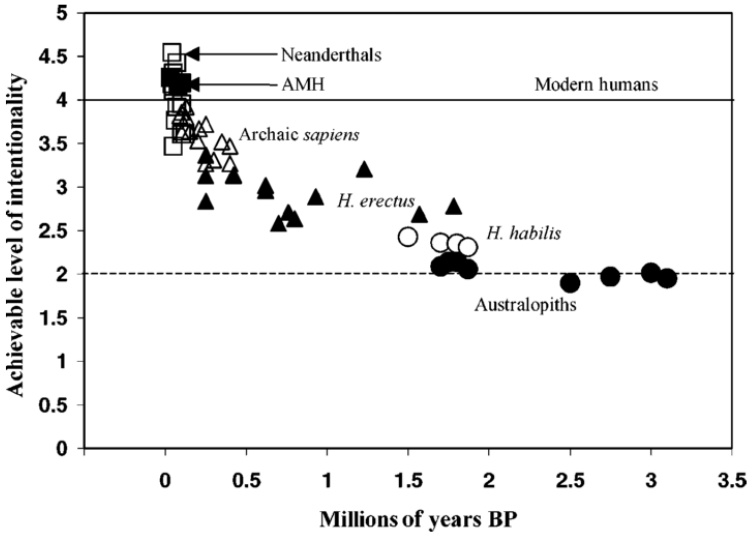


Figure 5 – AMH stands for "Anatomically Modern Humans"; BP for "Before Present" (source: Dunbar, 2003). Note that more recent insights suggest that Neanderthals could achieve four orders rather than five as indicated here (see Dunbar, 2014: 241-244).

Dunbar's fundamental assumption, then, is that brain size was the factor limiting how many levels of intentionality could be processed. Given that the archaeological record shows an increase in cranial volume in our lineage over the past 3-4 million years, this leads to the claim that our ancestors were able to handle increasingly more levels of intentionality across this period. Next, Dunbar argues that the maximum achievable level of intentionality put a limit on the sophistication of the language that could be developed at any particular stage, and thus the complexity of the activities that this language could support. For example, imagine that an early form of language existed in *Homo erectus* that was useful for coordination purposes in the here-and-now, but not sophisticated enough to support social bonding through gossip or telling stories. In that case, our ancestors at that time would not have been able to use this language for bonding (much) larger communities, as is assumed to have been the case in later stages (see Section 1.3.2 above). In this way, according to Dunbar's model, brain size limits achievable level of intentionality, which then limits sophistication of the language that can be developed, which in turn limits

the number of individuals that can be bonded in a coherent social group. Or phrased differently: increases in brain size over time "released" additional mindreading capabilities, which enabled more sophisticated forms of language, which in turn allowed for larger communities to be maintained in a coherent way.

As discussed in relation to the "third role" of language (Section 1.2.2 above), some researchers claim that it involves the capability to work at a handful of levels of intentionality to entertain human language "as we now know it", even when producing the most basic utterances (e.g. Sperber, 2000; Scott-Phillips, 2015). Combined with Dunbar's model set out above, this position necessarily entails that such language arrived late in our evolutionary history. After all, according to this model, the required ability to operate at such higher orders of intentionality was only available in anatomically modern humans. Earlier hominids may have had language, but this must then be assumed to have been of a lower degree of sophistication, given their limit at second- or third-order intentionality (for a discussion see Dunbar, 2014: chapter 7).

As said, in the chapters that follow I will first develop a perspective of economy, not so much looking at the limits of our mindreading capacity, but rather focussing on the *minimal* amount of mindreading needed for using language and dealing with stories. Chapter 6 then addresses the implications of this perspective for the practice of assessing mindreading experimentally. Finally, in the Conclusion I will return to the bigger picture set out in this first chapter and sketch the contours of how it should be updated in the light of the points developed throughout this thesis: after all, if the relationship between mindreading, language, and narrative is construed differently, this has potential consequences for the chronology of events assumed in the story of human evolution, and for the way these events are causally related.

Chapter 2

When narrative takes over: the representation of embedded mindstates in Shakespeare's *Othello*[*]

In recent times, researchers across a variety of disciplines in the humanities and social sciences have been interested in the human ability to process embedded mindstates, also known as 'multiple-order intentionality' (MOI): A *believes* that B *thinks* that C *intends* (etc.). This task is considered increasingly cognitively demanding with every order of embedding added. However, we argue that the way in which the information relevant to the task is represented in language (in particular, using a narrative) greatly influences how well people are able to deal with MOI cognitively. This effect can be illustrated by paraphrasing situations presented by a play such as Shakespeare's *Othello*: by the end of Act II the audience has to understand that Iago *intends* that Cassio *believes* that Desdemona *intends* that Othello *believes* that Cassio *did not intend* to disturb the peace. Formulated this way, using sentence embedding to express the intentional relationships, this is highly opaque. At the same time, we know that *Othello* has been understood and appreciated by innumerable different audiences for ages. What is it that the play's text does to make the audience understand all these embedded mindstates without undue cognitive strain? In this article we discuss six 'expository strategies' relevant to the representation of MOI and illustrate their working with examples from Shakespeare's *Othello*.

[*] This chapter was published as a separate article in *Language and Literature* (see Van Duijn, Sluiter, and Verhagen, 2015) and is inserted here without modification (with the only exception of a reference to this thesis on page 80, which is replaced by a footnote indicated with an asterisk). Since the article is co-authored, instead of "I" the first person plural is being used.

2.1 Introduction

Normally developed human adults are indubitably nature's mindreading champions: even when provided with only limited cues, we are often able to form strikingly elaborate understandings of what others think, believe, intend, desire, and so on. This capability is often referred to as 'Theory of Mind' or 'mindreading'. When multiple mindstates are embedded, the term 'multiple-order intentionality' (MOI) is used: A *knows* that B *believes* that C *intends* (etc.).

In recent years, scientists and scholars from various disciplinary backgrounds have made connections between MOI and literature. It is argued that readers of fiction not only have to keep track of what characters think (believe, intend, desire, etc.), but also of what they think these characters think that *other* characters think. Dunbar suggests that, in this way, an audience reading or watching Shakespeare's *Othello* has to work at higher orders of intentionality: 'they have to *believe* that Iago *intends* that Othello *imagines* that Desdemona *is in love* with Cassio' (2008: 414, italics added).[37]

This seems to pose a paradox: on the one hand, working at higher orders of intentionality is considered cognitively taxing (Kinderman et al., 1998; Stiller and Dunbar, 2007). On the other hand, it is clear that a play such as *Othello* has been understood and appreciated for ages, indicating that a normally gifted reader/watcher can follow it without undue cognitive strain. This suggests that some aspects of the exposition of information in *Othello* and comparable texts must alleviate the cognitive burden of processing MOI.

Using examples from *Othello*, we will discuss six strategies characteristic of (literary) narrative discourse that support the ability of readers to keep track of the mindstates of characters, even when these mindstates are mutually related and embedded in complicated ways. But before doing so, we will address MOI in some more detail, and discuss several possibilities to express embedded mindstates in language. We will conclude by discussing how our findings fit into a more general debate about the human ability to read minds.

[37] Similar examples drawn from different plays, popular culture, and novels can be found in, among others, Dennett (1987), Zunshine (2006) and Corballis (2011).

2.2 Mindreading, Theory of Mind, and Multiple-Order Intentionality

During the 1970s, several academic fields intensified their interest in how, and to what extent, humans and other primates can engage in taking the perspectives of others. Dennett (1971, 1987) introduced his concept of the 'intentional stance' and Premack and Woodruff (1978) made the term 'theory of mind' famous. Their work initiated a tradition of experimental and social research and raised fundamental debates among philosophers, psychologists, ethologists, neuroscientists, and contributors from other fields. A rich and insightful overview is given by Apperly in his recent monograph *Mindreaders* (2011). In this article we will follow Apperly's suggestion to drop the term 'theory of mind' (to avoid the implication that attributing mindstates is like having a theory) and refer to the set of mechanisms, routines, and tricks that humans apply to form understandings of other's mindstates as 'mindreading'.

In the past decade, several links have been pointed out between mindreading and literature. Some scholars have suggested that, when forming an understanding of the inner lives of fictional characters, we use the same mental capacities as when reading other people's minds in everyday social interaction (see e.g. Budelmann and Easterling, 2010; Cefalu, 2013; Palmer, 2004; Zunshine, 2006). Others have argued that reading fiction may train our mindreading abilities (e.g. Djikic et al., 2013; Kidd and Castano, 2013; Vermeule, 2010). This fits the broader idea that stories function as cognitive play: they may form a 'playground' where readers can develop various socio-cognitive skills without risking real-life social costs (Boyd, 2009). While the central focus of this article will be on embedded mindstates and the narrative expressed by Shakespeare's *Othello*, we will get back to these more general connections between mindreading and literature in our concluding section.

In various articles and books, Dennett (e.g. 1971, 1987) systematically works out the idea that mindreading has *recursive potential*: it can be self-embedded. In

line with the examples he discusses, this can be demonstrated as follows in Table 1:[38]

Table 1

P_1	[I know that] Will is a sailor.	*first order intentionality*
P_2	[I know that] Bill believes that Will is a sailor.	*second order intentionality*
P_3	[I know that] Mary believes that Bill believes that Will is a sailor.	third *order intentionality*
P_4	[I know that] Peter believes that Mary believes that Bill believes that Will is a sailor.	*fourth order intentionality*
P_5	[I know that] John believes that Peter believes that Mary believes that Bill believes that Will is a sailor.	*fifth order intentionality*
P_6	[I know that] Sally believes that John believes that Peter believes that Mary believes that Bill believes that Will is a sailor.	*sixth order intentionality*
P_n	[I know that] $Name_n$ believes that P_{n-1}	*n-th order intentionality*

The propositions in Table 1 show that the recursive procedure of embedding one proposition expressing a subject's mindstate into another, can produce propositions of any length and complexity. However, these propositions clearly become more opaque and their contents harder to grasp as the number of embeddings increases. This is presumably what led to the idea that there is a general upper limit to the number of orders of intentionality that humans can handle, as suggested by Dennett (1971) and tested experimentally by Kindermann et al. (1998) and Stiller and Dunbar (2007). They argue that the maximum for contemporary humans lies at around fifth order. In their experiments participants were presented with short narratives describing social situations that included several characters:

> Emma worked in a greengrocer's. She wanted to persuade her boss to give her an increase in wages. So she asked her friend Jenny, who was still at school, what she should say to the boss. "Tell him that the

[38] In this article we count orders of intentionality following Stiller and Dunbar (2007), who suggest that one should start from a first-person perspective by default (= first order). If this first-person perspective is not made explicit, it is added between brackets, as can be witnessed in Table 1. Note that fifth order in this article is the equivalent of fourth-order intentionality in Kinderman et al. (1998), who argue, accordingly, that the limit for contemporary humans lies at around fourth order.

> chemist near where you live wants you to work in his shop," Jenny suggested. "The boss won't want to lose you, so he will give you more money" she said. So when Emma went to see her boss that is what she told him. Her boss thought that Emma might be telling a lie, so he said he would think about it. Later, he went to the chemist's shop near Emma's house and asked the chemist whether he had offered a job to Emma. The chemist said he hadn't offered Emma a job. The next day the boss told Emma that he wouldn't give her an increase in wages, and she could take the job at the chemist's instead. (Stiller and Dunbar, 2007: 101-102)

After the narratives were read out to the participants, they were asked questions of increasing complexity, measured by the number of orders of intentionality as expressed through linguistic embeddings. They had to choose, for example, between 'Emma thought the boss believed that the chemist wanted her to work for him' or 'Emma thought the boss knew that the chemist had not offered her a job' (Stiller and Dunbar, 2007: 102). The level at which participants first failed answering such questions correctly was normally distributed with a peak around fifth order. However, it is not clear what exactly this means: the factor limiting performance at the higher levels was perhaps not so much the participants' ability to cognitively handle the *situation* presented by the narrative, nor was it the understanding of the *narrative* itself, but rather the participants' ability to process the multiply-embedded *sentences* of the questions. We suggest that the way mindreading tasks involving MOI are represented is crucial to the actual performance of the subjects facing these tasks.[39] In the next section we will take a closer look at the possible ways to represent MOI in language and narrative.

[39] This is also indicated by the fact that implicit mindreading tests, in which both the descriptions of the situations and the questions are presented in the form of 'narrative' movie clips, lead to increased success rates (see O'Grady et al., 2015). In addition, recent experimental findings by Carney, Wlodarski, and Dunbar (2014) indicate that genre impacts on people's understanding of MOI when reading narratives.

2.3 Representing embedded mindstates: from sentence to narrative

Within sentences, several linguistic devices are available for coordinating different perspectives (see, for example, Dancygier and Sweetser, 2012; Verhagen, 2005). At least in middle- and western-European languages, a central device is *complementation*, where a verb of cognition (*think*, *know*, *intend*, etc.) is specified by a complement clause, as in Table 1. However, if the representation of a MOI situation relies on such complex sentences only, it very soon becomes hard or even impossible for a reader or hearer to make the right inferences about the involved mindstates. Consider P_4 and P_5 in Table 1: these propositions are hard to link correctly to the situations referred to. This is in line with the fact that in corpora of narrative fiction and natural spoken discourse, sentences containing three or more embedded mindstates are very infrequent.[*]

At the same time it is widely accepted that we *do* regularly deal with higher-order mindreading in daily social life (Dunbar, 2003; Sperber, 1994), which suggests that there must be ways to communicate about this linguistically. In our example of Shakespeare's *Othello*, situations that require higher-order mindreading only emerge gradually in the course of the plot, after several characters have been introduced and several events have been narrated. Nowhere in the text does any single sentence express the embedding of more than three mindstates: only after a while does the situation give occasion to consider forming such sentences, summarising the state of affairs in terms of embedded mindstates. Dunbar provides such a sentence when he states that audiences watching a performance of *Othello* in a theatre have to work at fourth-order intentionality when 'they have to *believe* that Iago *intends* that Othello *imagines* that Desdemona *is in love* with Cassio' (Dunbar, 2008: 414;

[*] The Corpus of Spoken Dutch (CGN) features 10 million words, 1 million of which are fiction (mostly novels and short stories) read out aloud. Neither in the 9 million words recorded from spontaneous spoken discourse, business negotiations, radio reports, lectures, and so on, nor in the part recorded from fiction, can sentences be found containing four or more embedded clauses featuring a verb of cognition (think, know, believe, suspect, etc.) and/or communication (say, state, utter, bring forward, etc.). There is a handful (<5) examples of sentences containing three such clauses, which forms a stark contrast with the thousands of sentences containing one or two clauses featuring verbs of cognition and/or communication.

italics added). Since '*is in love* with Cassio' describes a mindstate without adding another embedding, for reasons of compatibility let us rephrase Dunbar's sentence as follows:

(1) [The audience *believe*] that Iago *intends* that Othello *imagines* that Desdemona *thinks* that Cassio is adorable.[40]

This fourth-order proposition does indeed summarise a crucial part of *Othello*'s plot: Iago, ensign to general Othello, determines on revenge after Cassio is appointed lieutenant over his head. He sets up an evil scheme in order to execute his revenge, which involves persuading the freshly married Othello that his bride, Desdemona, is unfaithful to him with the new lieutenant, Cassio.

However, we agree with Dunbar that it becomes more complicated than that. As Iago's scheming plan progresses, keeping track of 'who knows what' (including what they know that others know, etc.) involves reasoning up to at least sixth-order intentionality, following the same logic of counting. In the second act, Iago stage-manages a little riot. As a result, Cassio is discredited with Othello and suspended as a lieutenant. Next, Iago urges Cassio to ask Desdemona to plead with Othello for his (Cassio's) rehabilitation. In the meantime, Iago fills Othello with suspicion about what Desdemona's friendly stance towards Cassio means. The resulting situation is advantageous to Iago's plan: the more 'helpful' Desdemona is by pleading for Cassio, the more suspicious and jealous Othello becomes. With this situation on stage the audience would have to work at sixth order:

(2) [The audience *believe*] that Iago *intends* that Cassio *believes* that Desdemona *intends* that Othello *believes* that Cassio *did not intend* to disturb the peace.

This demonstrates that in the case of *Othello* it is possible to form sentences that in some respect offer faithful summaries of the contents of the plot, and at the same time contain up to six embedded mindstates. The text thus somehow

[40] The first order, 'the audience believe', is added between brackets to indicate that counting starts from that perspective, following Stiller and Dunbar (2007); see also our earlier Note 2.

represents (or prompts its readers to deal with) at least sixth-order intentionality. However, as stated earlier: nowhere does the text contain any sentences describing more than three mindstates at a time. This is where narrative takes over.

Dancygier's (2012) cognitive approach provides an insightful general framework for investigating how narrative texts produce meaning. In her view, texts provide *cues* that invite readers or hearers to make certain *inferences*. The elements that make up the story (events, character actions or traits, landscape's qualities, etc.) are 'compressed' (Dancygier, 2012: 59); that is: only *some* of the actual story's elements are represented by the text. For this representation, choices on several levels of language usage and narrative structure are made, that subsequently influence how readers or hearers 'unpack' the elements and construct the story in their minds.

In order to create what has been termed a 'fictional mind', a text provides cues that invite readers or watchers to imagine a (human) agent and attribute mindstates to him/her, using mostly the same mindreading abilities as they also use in actual social interaction.[41] Both lower-level linguistic choices (such as the use of a first-person versus a third-person pronoun in a sentence) and larger-scale choices on the level of narrative structure (such as telling in retrospect versus following the action as it unfolds) greatly influence the inferences made by readers, and thus the stories these readers construct in their minds (see Dancygier, 2012: 31–40).

In this way, using well-chosen cues, texts can prompt an audience to imagine a fictional social network of characters. As in real life, the audience can apply their mindreading skills to access intentional relationships between characters within this network if the context so requires. In cases such as *Othello* we argue that the process of imagining the right intentional relationships within the fictional social network of characters is *orchestrated* by structural properties of the text. These structural properties orchestrating the audience's

[41] This point is also made by Palmer (2004) and worked out for tragedy in Budelmann and Easterling (2010): inner lives and personalities of characters often remain largely unformulated in a direct sense, but the text 'model[s] the dynamics' of the reader's 'propensity to read minds' (Budelmann and Easterling, 2010: 290–292). For case studies dealing with mindreading and tragedy see Helms (2012); Sluiter, Corthals, Van Duijn et al. (2012).

mindreading process are what we will call *expository strategies* in the next section.

To summarise: anyone who knows the content of *Othello* would, upon consideration, admit that the propositions (1) and (2) in a certain way faithfully describe the intentional relationships between the audience and some of the characters. However, the text does not contain such complicated sentences. We argue that with the unfolding of the play's narrative, readers or hearers are invited to construct a fictional social network and approach it using mainly the same socio-cognitive skills as in real-life interaction. Narrative supports this process through several expository strategies. In the next section we distinguish six such strategies and discuss how they help the audience to construct the characters' intentional states and process the ways in which they are related and embedded.

2.4 The Expository Strategies of Narrative

The strategies we distinguish are: characterisation, focalisation/viewpoint alternation, framing, episodic structuring, time management, and redundancy. Their borders are not sharp: some overlap partly, others make use of related underlying mechanisms. None of them can, however, be fully reduced to any one of the others.

Before we discuss the strategies one by one, a few remarks have to be made about the fact that *Othello* is a play. The question of whether it is watched as a performance or read as a text is obviously important to the question of how the audience or readers are being supported in constructing mindstates of characters. Performances provide visual cues such as acted gestures, facial expressions, and proxemics (symbolic use of space; Hall, 1966) that probably trigger immediate responses in spectators. However, in what follows we will refrain from elaborating on the multi-modal experience of watching drama on stage, and restrict ourselves to analysing textual cues, given that *Othello* can be (and widely has been) understood and appreciated perfectly well by accessing its narrative through the reading of the text.

Some of our strategies build on established narratological theory, primarily intended for 'pure' narrative texts, such as novels or epic texts. Drama is different in that the interactions of the characters give the impression that they are unmediated. However, there is no doubt that the reader will come away with a clear sense of a story and a plot, of characters, setting, time, and place. Even though there is no overt narrator in drama, there still is a guiding hand, the theoretical instance situated between the actual author and the story told by the text, who is responsible for the textual choices that provide the reader with the cues to make the right inferences, and who directs their sympathies and antipathies.[42] All of this warrants the application of various narratological concepts in our following discussion of the strategies.[43] In addition, we will build on concepts and insights from cognitively oriented linguistics and literary studies, exploiting also their relevance to the processing of characters' mental states.

2.4.1 Characterisation

Clearly, before a reader can begin to read minds, agents have to be called into being to whom mindstates can be attributed. Only a few cues are needed to get this process started, after which numerous details can be added to adjust and deepen the picture drawn of a character and his or her inner life. Characterisation takes place at two separate though related levels: the *textual* level of the linguistic realisation and referencing of characters, and the *narrative* level at which the construction of characters as meaningful, 'human-like entities' (see Herman, 2009) is situated. Both are relevant to processing of mindstates.

At the textual level, in the case of a play (abbreviations of) the proper names of characters are commonly provided in the left hand margin to indicate who is speaking. Within character text, however, the range of linguistic options

[42] We will not go into those passages in which a character becomes a narrator of sorts, as in various soliloquies by Iago in which he comments on past or future events (an example is cited in Section 2.4.6), or in Othello's last lines, in which he instructs everyone on stage (and thereby also the audience) how to 'relate' the tragic events of the past hours to others (v. ii. 339–357); see also Maguire (2014: 1–4; 19–20).

[43] For narratology and drama in the context of our approach see Dancygier (2012: 146–164), and for a more general discussion see Hühn and Sommer (2009).

available for referencing other characters is endless: it includes proper names (e.g. 'Othello'), nicknames ('the Moor'), definite or indefinite descriptions ('the general', 'a man'), pronouns ('he'), and more. For the purpose of this article two aspects are especially relevant. Firstly, and rather obviously, proper names (or nicknames) can form unique labels that support the readers' memories in keeping different agents and their mindstates apart: 'A person intends that another person believes that again another person loves yet another person', is (even) harder to grasp than the statement that 'Iago intends that Othello believes that Cassio loves Desdemona', whether or not these names had been introduced before. Secondly, and more importantly, proper names can activate so-called 'rich frames' (Dancygier, 2011). A rich frame is a bundle of rather specific background knowledge present in readers, structured such that if one aspect is mentioned, a set of other relevant aspects are also activated. Anyone mentioning 'Mozart' refers to a particular historical person, but also implicitly activates aspects such as 'composer', 'genius', 'fame', '18th-century classical music', 'early death', 'man', and so on. In the case of fictional characters *created* by a text (rather than 'reincarnated' from history or existing stories and myths), the background knowledge contained in the frame associated with their name will clearly be much more rudimentary initially (cueing only, for instance, 'man' or 'foreign'), but as the narrative develops the frame can be enriched very quickly (see also Framing in Section 2.4.3).

This is where the second level of characterisation comes in, the narrative construction of characters. As soon as the involved minds are 'labeled', attributes can accumulate. For example, the readers' pictures of Othello and Desdemona will be modified and enriched as they learn in the course of the first act that these two are a respected general with an 'exotic' background and the daughter of a Venetian aristocrat, who got married in secret. With every detail added, each new mention of Othello or Desdemona will activate a richer background knowledge, which in turn enables more profound understanding of the way they think, what their plans are, how they will react, and so on.[44]

[44] In prints (both ancient and modern), the text is often preceded by a list summing up the *dramatis personae*, followed by a brief description, such as 'Roderigo, a Venetian Gentleman' (cf. Ridley, 1965: 2). In this way, this form of characterisation can sometimes even begin before the play itself begins.

Both narrative characterisation and referencing of characters at the textual level are crucial in enhancing *accessibility*, the ease (or difficulty) with which information relevant to successful processing of meaning can be retrieved from memory. In line with Sperber and Wilson's (1986) 'Relevance' principle, Ariel (1990: 165–168) points out that the amount of processing power needed to interpret a linguistic utterance is inversely proportional to the availability of a context in which this utterance makes good sense. Lower accessibility means thus that more processing power will be invested in the search for such a context. Ariel argues that some patterns of referential expression usage focus this search process, so that the reader will need to invest less processing power in order to come to a relevant interpretation. In the example of 'A person intends that another person believes that again another person loves yet another person', and also in the sentences (1) and (2) from Section 2.3, a handful of agents are prompted, as it were, out of the blue. By contrast, in a concisely constructed piece such as Shakespeare's *Othello* this prompting is both more gradual and more redundant: characters appear on stage and interact, they leave, come back again in other scenes, are addressed by other characters when present, or referenced when absent, etcetera. As a consequence, the core set of characters will attain an increasing degree of presence in the short-term and working memory of the readers, or in Ariel's words: their accessibility increases, hence the amount of processing power needed for interpretation when they figure in a linguistic expression goes down (1990: 11–31; see also Vandelanotte, 2012: 203–207, and our Section 2.3.6 on Redundancy).

Finally, some details of character construction are provided by another character, and may therefore be as illuminating about the speaker as about the object of his speech. Consider, for instance, the introduction of Cassio: he is first presented as an 'arithmetician', a 'Florentine', but also as a 'bookish theoretic' whose soldiership is 'mere prattle without practice' (I. i. 19–26).[45] Although especially the first attributes may seem objective facts, one must be watchful: everything the readers learn here comes from the mouth of Iago, and must thus be considered in the light of his specific interests (see also Focalisation/

[45] The used edition of *Othello* is Ridley (1965), also known as 'Arden 2'. References are to act, scene, and lines (act. scene. line) of this edition.

viewpoint alternation in the next section). The clearly negative valuation of the details he provides reveal Iago's distaste for Cassio, and thereby provide the reader with another basis for following Iago's wily plans.

In this way, with characters talking to and about each other, readers acquire knowledge of their relationships and stances towards one another; in other words, the readers gain increasing insight in the fictional social network. As in real life, they may start off with only some stereotypical expectations, which are then reinforced, readjusted, or weakened with every new detail mentioned. Against this crucial background of characterisation, more specific character mindstates can eventually be constructed and understood.

2.4.2 Focalisation and viewpoint alternation

Characters are not just fictional agents allowing the attribution of mindstates, but they also *perceive* the events that constitute the narrative's plot from their point of view. The text orchestrates the reader's inclination to align with the perspective of different characters, and to perceive these events 'through their eyes'. In narratology this is referred to as *focalisation* (Bal, 2009; De Jong, 1997: 313–319), but we take the concept more broadly in this article and extend our interest to the more general usage of all kinds of linguistic cues that mark viewpoint shifts. Elaborating further on Cassio's introduction as an 'arithmetician' and a 'bookish theoretic', it is clear that Iago, who presents this information, is spinning it: his character text prompts the readers at least temporarily to look at a part of the narrated world (in this case, the character Cassio) from his (Iago's) particular viewpoint. This viewpoint is tinged by his jealousy rooted in having been passed over for the military promotion which Cassio received in his place. Through such details provided from Iago's viewpoint, the readers gradually become familiar with his anger. This puts them in the ideal position to correctly interpret his performance when he is asked to explain to Othello the very messy situation in which an inebriated Cassio stabbed Montano. The reader knows that the whole riot leading to this unfortunate event had been set up by Iago himself, but Othello is of course unaware of this when he turns to Iago for enlightenment:

Iago Touch me not so near,
I had rather ha' this tongue cut from my mouth,
Than it should do harm to Michael Cassio:
Yet I persuade myself, to speak the truth
215 Shall nothing wrong him. Thus it is, general:
Montano and myself being in speech,
There comes a fellow, crying out for help,
And Cassio following him with determin'd sword,
To execute upon him . . .
(ii. iii. 211–219)

Iago's wily first three lines make him out to be a steadfast friend of Cassio's, one who would rather suffer himself than see his friend come to harm through his words. But surely, Iago puts forward, if he keeps to the 'truth' he will not harm his friend. At this point the readers are in a perfect position to grasp Iago's true attitude and evil intentions: they can take both his point of view and that of the guileless Othello to judge the effect of the lines that follow. Iago's version of the events is 'correct', yet it leads to an exaggerated (as the reader knows) picture of reprehensible guilt of Cassio.

Note that Iago's words include a representation of Cassio's viewpoint (who is 'following [Roderigo] with determin'd sword') – a case of 'embedded focalisation'. Here the text urges readers to do a 'double take' on the event witnessed earlier, but then through the eyes of Iago, and *via him* again through the eyes of Cassio. Through these viewpoint shifts the readers can conclude that Othello, Montano, and Cassio, who were not present when Iago discussed the set-up for the riot with Roderigo, now have a false impression of what just happened, shaped by Iago's sly account (see also Maguire, 2014: 35–38).

In this way, throughout the play the readers take a whole carousel of different viewpoints: they do not have to consider mindstates from the 'outside', starting from one fixed perspective, as is the case in multiple-order propositions such as (1) and (2) given earlier, but the text makes them experience these viewpoints, as it were, 'from within'.

2.4.3 Framing

Specific details can have a great influence on the impression a reader has of the characters and their mutual relationships. Such details can activate knowledge already present in readers, both general knowledge about the world (for example the scenario of 'adulterous lady provokes jealous husband') and more specific background knowledge about elements of a story (for example: 'Othello is a Moor, so the Venetian aristocrats probably consider him an outsider'). Here, the concept of *framing* is relevant (Coulson, 2001; Fillmore, 1976). Frames bring along a certain *topology*, detailing how informational cues in the new context should be connected and interpreted. Knowledge of particular frames is cultural common ground in readers' minds. If a frame concerns the behaviour of humans, its topology usually contains 'stereotypical roles', information about what characters are likely to do, think, believe, intend, etcetera.

In the course of Shakespeare's play, the reader is provided with the information that Iago, out of *revenge*, sets up *a wily plot* by making Othello think that his wife, Desdemona, is *adulterous*. These terms all link to well-known frames that come with stereotypical roles, including particular mindstates. For example, in the case of adultery: 'A and B are married'; 'A and C are in love'; and 'A and C do not want B to know that A and C are in love'. By mapping the characters on stage onto the structure of this topology, the reader can make the relevant inferences about intentional relationships without having to process all mindstates from scratch.

Given the benefits of spectator-sight, readers can also witness how one character uses framing to influence another character's mindstate. Consider again the step in Iago's scheme where he *intends* Othello to *believe* that Cassio *is in love* with Desdemona and that this love is reciprocated (cf. (1) above). The readers witness how Iago attempts to *reframe* the friendship between Desdemona and Cassio as adultery in Othello's mind (cf. 'frame-shifting': Coulson, 2001). As a result, Othello ascribes different (and false) intentions to Desdemona and Cassio when he watches their contacts and actions. Again, readers do not face the task of processing step by step a completely novel situation: using existing knowledge of the adultery and friendship frames, they can grasp in a more holistic way what is going on in Othello's mind.

2.4.4 Episodic structuring

When reading the play, it is clear that it is divided into several more or less bounded episodes. Within such an episode, a relationship between characters can be elaborated, a character can be struck by an idea, change his mind, or acquire an insight. Typically, one episode includes only a limited number of characters: it is mostly either a monologue (such as those given by Iago), a dialogue, or a discussion between a few characters (see also Stiller et al., 2003). The crucial point is that the information provided in any episode can, as it were, be 'collapsed' and conveniently taken to subsequent parts of the play. Here it can be referred to, and embedded into a new structure.

Stories routinely embed earlier episodes by circumscribing them pragmatically, using a key feature of their content or presentation. For example, in the first act Roderigo and Iago spend quite some time discussing why they hate Othello. Later, when Iago needs to convince Roderigo to cooperate (again) in one of the steps of his scheming plan, he only needs to refer to this discussion to reactivate the chain of shared and individual negative experiences with Othello in both Roderigo's and the readers' minds. In this way, the play can carefully build up a complex plot by first presenting several episodes that are easy to understand, and then embedding these episodes into one another by calling them up later.[46]

To summarise: the fact that the story develops episodically, with each episode updating the mindstates of only a limited subset of the total number of characters, makes it possible that a complex structure emerges in a manageable way.

2.4.5 Time management

Story time can be managed flexibly through flashbacks or anticipations. Information need not be presented in the actual order of events 'in the world': this is the narratological distinction between the order in which events 'really' happened (known as the *fabula*) and the order in which they are being told

[46] See also Dancygier's (2012) concept of *narrative anchors*.

(*story*). In the fabula of Shakespeare's play, for example, Othello and Desdemona get married before the dispute between Iago and Roderigo with which the play opens. However, in the play's story it takes about 200 lines starting from the dispute before the marriage is narrated as a past event. This principle not only opens a possibility for redundant narration (as will be discussed later) but also for calling in past events for clarification or further elaboration.

Events that take years may take only seconds to narrate, and an instantaneous event may be described at extravagant length (see e.g. Schneider, 2003). In this connection some scholars have proposed that Othello contains a 'double time scheme' (Ridley, 1965: lxvii–lxx). On the one hand, the main action moves fast: there is a night, a day, another night, and another day on the evening of which the play ends with the killing of Desdemona. This high pace greatly adds to the overall suspense, as well as to the credibility of Iago's scheme. After all, had there been plenty of time between all of his cunning moves, Othello, Cassio, Desdemona, and Roderigo would have been able to speak to each other, and figure out Iago's evil intentions. At the same time, however, the rapid movement of the action 'makes nonsense of the whole business', as Ridley (1965: lxix) puts it, leaving no time during which the supposed adultery of Cassio and Desdemona could have occurred. Ridley concludes by noting that 'what Shakespeare is doing is to present, before our eyes, an unbroken series of events happening in "short time", but to present them against a background of events not presented but implied, which gives the needed impression of "long time" … He knew to a fraction of an inch how far he could go in playing a trick on his audience' (1965: lxx). What Ridley calls a 'trick', may also be regarded as managing the story time in such a way that the reader is supported in making the right inferences about who-knows-what for the plot to work.

2.4.6 Redundancy

Finally, there is the aspect of redundancy. Some information is simply provided more than once in the course of the play. Mostly, the repeated information is

incorporated somewhat differently, using stylistic variations, shifting points of view, or flashbacks and flash-forwards.

In *Othello*, examples of this can be found in the various soliloquies delivered by Iago. Here he ruminates on what has just occurred and on what is about to occur next. In effect, he rehearses important pieces of information from past and future events for the audience. In fact, this also effects the emergence of a form of drama-internal 'narrating voice', offering reflection on the ongoing action (see also Dancygier, 2012: 147–148; Maguire, 2014: 1–4). An example can be found at the end of Act II (II. iii. 305–353): because of the riot that Iago has set up with the help of Roderigo, Cassio has just been suspended as a lieutenant. However, he has no idea that Iago is in fact the evil genius behind his suspension. In a dialogue between the two men, Iago expresses his faked compassion with Cassio, and provides him with a piece of advice on how to regain his position: he should seek the intercession of Desdemona. It starts as follows:

Iago

305 I'll tell you what you shall do... Our general's
Wife is now the general . . .

309 Confess yourself freely to her, importune
Her she'll help to put you in your place again . . .

Cassio

317 You advise me well.

After this devious advice has been given, Cassio repeats it to himself:

Cassio

320 I bethink it freely, and betimes in the morning will I
Beseech the virtuous Desdemona, to undertake for
Me; I am desperate for my fortunes, if they check me
here.

Then Cassio leaves and Iago remains on stage alone. He begins a monologue, in which he first looks back on the advice he just gave to Cassio:

Iago

And what's he then, that says I play the villain,
When this advice is free I give, and honest,
Probal to thinking*, and indeed the course — * "*a wise advise*'
To win the Moor* again? — * *i.e. Othello*

How could his advice be considered evil? It is obviously a good way to placate Othello. Of course both he and the reader know better. Iago predicts that Desdemona will be easy to convince:

For 'tis most easy
The inclining* Desdemona to subdue**, — * "*compliant*' **"*convince*'

Next, Iago predicts that Othello will be susceptible to Desdemona's pleas: in fact, she could get him to do anything at all:

and then for her
To win the Moor, were 't to renounce his baptism,
All seals and symbols of redeemed sin,
His soul is so infetter'd* to her love, — * *from "fetter", i.e. "chain"*
That she may make, unmake, do what she list*, — * "*as she thinks fit*"
Even as her appetite* shall play the god — * *"his desire for her"* ("her is objective, not possessive", Ridley, 1965: 87)
With his weak function.

Then Iago looks ahead to what the next step in his own scheming plan is going to be: he will fill Othello with suspicion about the relationship between Cassio and Desdemona. Meanwhile, he *again* mentions both the advice he gave to Cassio and Desdemona's presumed willingness to go and plead with Othello. And this time it is clear that he is entirely and overtly conscious of the evil nature of his scheme:

How am I then a villain,
To counsel Cassio to this parallel course,
Directly to his good? Divinity of hell!
When devils will their blackest sins put on,

They do suggest* at first with heavenly shows — * "*seduce*"
As I do now: for while this honest fool
Plies Desdemona to repair his fortunes,
And she for him pleads strongly to the Moor,
I'll pour this pestilence* into his ear — * *the "pest" of suspicion*

Finally, Iago tells the reader what he thinks the result will be: the more Desdemona argues in Cassio's favour, the more suspicious Othello will get. This is, obviously, good from the perspective of Iago's true interests and intentions, as it brings him closer to his revenge:

And by how much she strives to do him good,
She shall undo her credit with the Moor;
So will I turn her virtue into pitch* — * "*tar*"
And out of her own goodness make the net
That shall enmesh* 'em all. — * "*trap*"

This is where Iago's monologue ends. In what follows, Cassio acts upon Iago's advice and asks Desdemona to plead in his favor. Othello turns out to be amenable to this, and Iago indeed manages to fill him with suspicion about why Desdemona is so eager to plead for Cassio's rehabilitation.

Notice that the first of these steps (Cassio asking Desdemona to plead with Othello for his rehabilitation) is presented to the reader as much as five times: once when Iago first gives the advice to Cassio, once when Cassio rephrases the advice for himself, twice in Iago's monologue, and once again when Cassio actually asks Desdemona to go and plead.[47]

It is clear that *repetition* of cues from which readers can construct the intentional relationships important for the development of the play's plot, increases the ease with which these relationships can be processed cognitively.

[47] An acute observer may even find a sixth time in ii. iii. 376–377.

2.5 Conclusions and Discussion

2.5.1 Viewpoint all the way down

We suggested that sentence embedding is well suited for handling two, perhaps three orders of intentionality at a time, but usually not more. For higher orders, narrative takes over. Indeed, Shakespeare's *Othello* uses lower-order sentences to build a narrative that represents the higher-order situations constituting the plot. We argued that narratives such as the one presented in *Othello* provide support and scaffolding for readers' abilities to process multiple-order intentionality by providing cues that prompt them to construct a fictional social network using mainly the same socio-cognitive skills as in real-life interaction. Structural properties of the narrative, here termed expository strategies, orchestrate step by step how the network of intentional relations relevant to understanding the plot should be constructed. We distinguished six such expository strategies: characterisation, focalisation/viewpoint alternation, framing, episodic structuring, time management, and redundancy. All of these can serve the purpose of *viewpoint management*: they impact the text's capability to orchestrate how the reader should navigate through a complex network of embedded and interlinked viewpoints. Whereas this is obvious and well established throughout the literature for the strategy of focalisation/viewpoint alternation, the other strategies have not so far been systematically linked to viewpoint management. Our approach shows how the issue of managing viewpoints pervades in virtually all levels of the text and cuts through the borders of traditionally distinguished categories in narrative theory. This observation also reflects recent developments in the study of language and literature: throughout the past decade scholars coming from various angles and backgrounds have arrived at approaches in which viewpoint (including its cognitive pendant: the mindstate) is the central issue (for a recent overview see Dancygier and Sweetser, 2012; see also Bernaerts et al., 2013; Hühn, Schmid and Schönert, 2009; McConachie and Hart, 2006; Palmer, 2004; Vandelanotte, 2009).

2.5.2 Context and directions

Our analysis fits into several broader debates. There is increasing evidence that the human mindreading capability has (at least) two layers that function and develop to a certain extent independently: the first layer is mostly in place from birth and is shared with other species in nature; on top of that there are one or more layers that are uniquely human and develop over a longer period of time (Apperly and Butterfill, 2009). The first layer is mostly associated with 'sensing' feelings, directedness, and intentions of others in an immediate, automatic, and irresistible fashion. The second layer is associated with developing a more sophisticated understanding of thoughts and motives, and with appreciating such communicative leaps as irony and sarcasm. It can take up to the age of 12 or 14 before an adult level of understanding other minds has been achieved (Papp, 2006). It has been suggested that the development of the second layer is to a great extent influenced by social interaction and increasing experience with communicative situations, but also by language acquisition as such. Evans and Levinson (2009), for example, conclude that by learning to master the grammatical practice of sentence embedding, children implicitly improve their cognitive skills for interpreting daily-life situations, particularly those involving embedded mindstates. Experiments by Lohmann and Tomasello (2003) further support this: children who are trained in perspective-shifting discourse containing embedded sentences, perform better at mindreading tests than their peers who have been trained in different forms of language use (see also De Villiers and De Villiers, 2003; Milligan et al., 2007). Just as the acquisition of grammar supports the development of social cognitive skills, it is perfectly thinkable that learning to understand and produce *narratives* also improves a child's advanced mindreading abilities. This is a promising direction for future research (see also Gallagher and Hutto, 2008).

In addition, the working of our six strategies can be tested experimentally by composing several versions of narrative excerpts, manipulating one strategy at a time, and measuring how well readers are able to process the multiply-embedded mindstates. We also suggest that an empirical approach focused on development and language acquisition would be highly viable, investigating the potential link between the acquisition of the linguistic and narrative

features comprised in the six expository strategies and the development of mindreading abilities in children and adolescents. Whereas existing research in this domain has focused on the developmental step from no or very basic understanding of intentionality at age 1–2, to the full appreciation of (false) belief-states at age 4–5 (Milligan et al., 2007), it would be of particular interest to trace the steps beyond that point: those involving higher-order mindreading.

In this article, we focused on the possibility of employing insights from rich traditions in the humanities, such as narratology and the study of grammar and semantics, to contribute to current debates concerning human cognition. This demonstrates how research on such multifaceted topics as mindreading can benefit from the integration of insights and methods from different academic disciplines across both the sciences and the humanities – an example of how EO Wilson's idea of *consilience* (1998) can be brought into practice.

Chapter 3

Not afraid of Virginia Woolf: embedding and polyphony in the novel[*]

Cognitive literary scholar Lisa Zunshine argues in her foundational 2006 book *Why we read fiction* that a certain class of literary novels push readers to the limits of their cognitive abilities. She suggests that this might be why some of us "are afraid of Virginia Woolf": when reading for example *Mrs Dalloway*, according to Zunshine we have to process complexly embedded mindstates of the type "A *suspects* that B *knows* that C *wants* (etc.)". I agree that Woolf's fiction (and *Mrs Dalloway* in particular) is a great object of study for anyone interested in the complexity posed by representing a wealth of different mindstates in one story. However, in this chapter I question whether this complexity is adequately conceptualised as a series of embedded layers. I revisit the excerpt from Woolf's *Mrs Dalloway* used by Zunshine to support her claim and analyse the linguistic and narrative "cues" that prompt the reader to imagine a rich "thoughtscape" of character mindstates, which are shown to be mutually related and interlinked in a variety of ways rather than being just embedded. This leads not only to a different view on the excerpt from Woolf's novel, but also has important implications for how the general relationship between discourse structure and cognitive processing should be formulated.

[*] Chapters 2, 3, and 4 were written as independent articles; see also the Reading Guide above.

3.1 Introduction

Humans have the capability to form beliefs about the mind-states of others; this is generally referred to as "mindreading", "mentalising", or having a "theory of mind". Since others have such beliefs too, in social contexts involving several human actors, it has been claimed that beliefs about mind-states can become embedded into one another: A *thinks* that B *intends* that C *believes* [etc.] that X. This is referred to as "multiple-order intentionality". This chapter focuses on the representation of multiple, interconnected mindstates in novels, addressing in particular the example used by cognitive literary scholar Lisa Zunshine (2006; 2012): a passage from Virginia Woolf's *Mrs Dalloway* (1925).

The capability to take others' perspectives is found in some form in various species of mammals and birds.[48] Especially our close relatives in nature, most notably chimpanzees and bonobos, seem to be capable of assessing what others around them can see and know, and to some extent even what their goals and intentions are. However, when it comes to mindreading, humans are indubitably nature's champions. Not only are we the only species that can reliably handle so-called "false beliefs", a landmark passed by any normally developing human individual from around four years of age, but we are also the only ones capable of dealing with tasks that feature multiple, mutually related mindstates at the same time (Dennett, 1987; Dunbar, 2003). Imagine, for example, someone organising a surprise party for her brother. She calls him, pretending that she needs his help with something only he can do next Friday evening, which happens to be his birthday. She has to make sure that he *does not understand* that she *intends* to invite all his friends to her house. Also, she has to make her brother's friends *understand* that her brother *does not know* about the party, and she might have to *anticipate* that he *may ask* his friends whether they *know* that he will not be available on the night of his birthday, and so on.

This shows that mindreading complexity easily adds up even in quite ordinary situations of everyday social life. Researchers have proposed to

[48] For an overview see Apperly (2011: 46-56). See e.g. Call & Tomasello (2008) and Yamamoto et al. (2012) for perspective-taking in great apes, De Waal (2013) for a broader view on this skill in mammals, and Clayton et al. (2007) for perspective-taking in crows. See also Heyes (2012) for a critical review of some of the evidence.

measure such complexity in terms of the number of "layers" or "orders of intentionality" involved. Reasoning about what someone else is thinking comprises two orders, reasoning about, for example, what someone else wants you to understand involves three, reasoning about what someone else wants you to think about another's thoughts involves four, and so on (for a discussion see Chapter 1, Section 1.2; Dennett, 1987: chapter 7). It has been argued that all forms of mindreading observed in non-human animals can be characterised in terms of two such orders at the maximum—their limit is situated at or below "second-order intentionality". In contrast, humans are capable of handling tasks involving three or more orders, at least from a certain age onwards (see Henzi et al., 2007). Yet human mindreading sophistication is also not unlimited: experiments suggest that the majority of normally developed adults lose track after around five orders of intentionality (see Chapter 1, Section 1.2; Kinderman, Dunbar, and Bentall, 1998; Stiller & Dunbar, 2007; Launay et al., 2015).

Dunbar (2005; 2008) has suggested that this limit on the number of orders of intentionality we can typically deal with, in turn constrains the cultural practice of producing narratives: stories are likely to involve up to around fifth order, but not more, due to the cognitive limits on readers' and writers' abilities to understand and write such stories. Dunbar also argues that people take a certain delight in being pushed to the boundaries of their cognitive limits, and that it may be a hallmark of literature to do so. Literary scholars, as well as philosophers, psychologists and other cognitive scientists, have indeed used excerpts from literary texts to demonstrate how complex and multi-layered mindreading can get. Arguably the most famous example is Zunshine's analysis of the passage from Woolf's *Mrs Dalloway* in which Hugh Whitbread, with the aid of Richard Dalloway, writes a letter to the Times on behalf of Lady Bruton. According to Zunshine's analysis, readers of the novel at this point have to understand that "Woolf *intends us to recognize* [...] that Richard *is aware* that Hugh *wants* Lady Bruton and Richard to *think* that because the makers of the pen *believe* that it will never wear out the editors of the *Times* will *respect* and publish the ideas recorded by this pen" (2012: 207, italics in original). Zunshine

labels this task as "sixth-level" intentionality (see also Section 3.3 below).[49] On this basis she suggests that "certain aspects of Woolf's prose do place extraordinarily high demands on our mind-reading ability" (203-204), which, on the one hand, could explain why some of us are "afraid of *Mrs. Dalloway*" (202), while, on the other hand, she suggests that it may explain why we read such demanding fiction at all: we take a certain delight in pushing our mindreading abilities to their limits and putting them to the test.

In this chapter I suggest to turn the issue around: instead of asking how many orders humans can maximally handle and how this may affect or constrain the stories that can be produced and understood, I ask what readers minimally need in order to deal with a passage such as the one from Woolf's novel that has been claimed to require such a high degree of mindreading sophistication. The focus is thus not on the presumed *limits* of readers' capability to cognitively process mindreading, but rather on the *economy* and *expediency* of this capability. Using and extending concepts from cognitive linguistics/stylistics and narratology (building especially on Dancygier, 2012), I will argue that the structure and language of Woolf's novel prompt a rich *thoughtscape*, a network of interlinked and/or embedded mindstates, and at the same time support the reader in processing this thoughtscape in a natural way. On the basis of this I will reconsider the way in which the relationship between discourse structure and cognitive processing should be conceptualised.

3.2 Mindstates in literature

3.2.1 An early cognitive literary scholar

Virginia Woolf's *Mrs Dalloway* (1925) features a day in the life of Clarissa Dalloway, a woman from the London high society in the years after the First World War. The story begins with her walking through the city to buy flowers,

[49] Zunshine's analysis was originally published in her 2006 book *Why we read fiction* and then reused in 2012. Other examples of such analyses can be found in Dunbar (2008), Corballis (2011), both citing Shakespeare (*Othello* and *Twelfth Night*, respectively), Palmer (2012), and Dennett (1987), all in similar ways. See Chapter 2 above for more examples and for a discussion.

in preparation of one of her regular, fashionable parties taking place that evening. Much of the day is spent thinking about choices she has made in her life and people she has met. She encounters some of these people while wandering around or preparing the house and eventually sees most of them at the party. For an important part, thoughts of the past are triggered by a remarkable encounter between Clarissa Dalloway and Peter Walsh, a lover from her youth, who disappeared to India decades ago after she had declined his wedding proposal. From his letters she knew he was planning to come back to England one of these days, but his sudden visit greatly surprises her.

As Peter Walsh leaves the house, point of view stays with him for a while, after which it shifts to various other settings and events somehow related to Clarissa and her party. Narration in the novel is thus also "wandering", in the sense that the story is related by an omniscient narrator who constantly shifts from the perspective of one character to another, across different settings in time and place as well as within one setting. As a result, readers benefit from what could be called a "360-degree view" of the represented situations and events. In this way, we are enabled to build an increasingly profound understanding of all kinds of relationships between the characters, including shared knowledge of the past (and possibly different perspectives on things that happened), aspects of personalities (and different perspectives on these aspects), affections, intrigues, worries about what others may think of oneself, worries about what others may think that one thinks about them, and so on.

All of this warrants that the novel is an excellent choice for an analysis of mindreading and mindstates in fiction. This has not gone unnoticed by scholars interested in cognitive literary analysis: Mrs Dalloway is a popular guest at *their* parties (see, for example, Dancygier, 2012; Oatley, 2011; Vermeule, 2010; Vandelanotte, 2009). However, these scholars are not the first to unveil the relation between Woolf's work (or literature more widely) and the representation of the workings and states of the human mind. It is a characteristic of modernist fiction more generally to be fascinated by people's inner lives (Korsten, 2005: 227-249) and when reading what Virginia Woolf had to say about this herself one is easily tempted to call her an early cognitive literary scholar. Consider the following passage from an essay, which she originally read as a lecture in Cambridge in 1924:

> My first assertion is one that I think you will grant—that every one in this room is a judge of character. Indeed it would be impossible to live for a year without disaster unless one practised character-reading and had some skill in the art. Our marriages, our friendships depend on it; our business largely depends on it; every day questions arise which can only be solved by its help.
>
> [...] But it is the art of the young. In middle age and in old age the art is practised mostly for its uses, and friendships and other adventures and experiments in the art of reading character are seldom made. But novelists differ from the rest of the world because they do not cease to be interested in character when they have learnt enough about it for practical purposes. They go a step further; they feel that there is something permanently interesting in character in itself. When all the practical business of life has been discharged, there is something about people which continues to seem to them of overwhelming importance [...] And this I find is very difficult to explain: [...] what the impulse is that urges them so powerfully every now and then to embody their view in writing.
>
> [...] I believe that all novels, that is to say, deal with character, and that it is to express character—not to preach doctrines, sing songs, or celebrate the glories of the British Empire, that the form of the novel, so clumsy, verbose, and undramatic, so rich, elastic, and alive, has been evolved. (Woolf, 1924: 2-11)

By just replacing a few instances of the term "character" by the terms "mind" or "mindreading", one can read a doctrine fitting seamlessly in the theories of various present-day cognitive scientists and scholars. Woolf would be taking a position close to Hutto's "narrative practice hypothesis", stating that every human being is in fact a teller of folk-psychological tales in the practice of daily social interaction, whereas some individuals take this skill beyond routine and make a living out of it (Hutto, 2008; see also Chapter 1, Section 1.1.4). She would line up with cognitive literary scholars such as Palmer (2004) and Herman (2009), whose work gives expression to the idea that the construction of human "minds" is the central feature of fiction. Virginia Woolf would probably agree that literature can be seen as an environment for experimenting, or even as a

training ground for practicing how to assess other's inner lives, as argued by Zunshine (2006), Boyd (2009); Vermeule (2010), and Oatley (2011, who also cites parts of Woolf's 1924 lecture). And she would be fascinated, but not in the least part surprised, by the results from controlled experiments showing that reading fiction can enhance performance on various sorts of mindreading tests (e.g. Kidd and Castano, 2014; Djikic et al., 2013).

3.2.2 Zunshine's "sociocognitive complexity"

Considering all this, I argue with Zunshine that the passage from *Mrs Dalloway* about the writing of a letter to the *Times* suits an analysis in terms of mindreading well. However, in what follows I will question three assumptions Zunshine makes in her approach: (i) that this passage (or the novel more generally) confronts readers with a form of complexity adequately conceptualised by counting layers of embedded mindstates; (ii) that these embedded layers pose a highly demanding cognitive processing task at the limit of these readers' abilities; and (iii) that their appreciation of the literary work is somehow affected by this layered complexity. With regard to the targeted passage (cited at the beginning of this chapter) she writes:

> To grasp the full meaning of this passage . . . we first have to process several sequences that embed at least five levels of intentionality. Moreover, we have to do it on the spot, unaided by pen and paper and not forewarned that the number of levels of intentionality that we are about to encounter is considered by cognitive scientists to create "a very significant load on most people's cognitive abilities" (Zunshine, 2012: 207)

The analysis of *Mrs Dalloway* was first published in Zunshine's 2006 monograph *Why We Read Fiction: Theory of Mind and the Novel* and repeated in the 2012 *Introduction to Cognitive Cultural Studies* of which she is the editor (2012: 193-213). She builds on it in more recent work, in which she introduces the term "sociocognitive complexity" to refer to "patterns of embedment of mental states

within mental states in fiction" (2012a: 13).[50] She suggests that different genres may implicitly expect different levels of "sociocognitive literacy" (i.e. aptitude to deal with embedded mental states) from their readers: for example, Jane Austen's *Pride and Prejudice* may in its original form expect readers to deal with fourth- or fifth-order intentionality, whereas the same story in a comic-book adaptation or study-guide synopsis might be "downgraded" to second order, thus anticipating lower sociocognitive literacy as appropriate in these genres (2012a: 17). In addition, in several publications Zunshine links the levels of intentionality to perceived literary quality, arguing that there may be a "literary sweetspot" at which the number of embedded layers is optimal: given a particular reader's sociocognitive literacy there can be too few embeddings, taking the challenge away, or too many, posing too high a cognitive load, both leading to lower appreciation of a text (for an experimental approach see Whalen, Zunshine, and Holquist, 2012; for more discussion see Zunshine 2012; 2012a; 2011; see also Dunbar, 2005).

Zunshine's approach of analysing literary texts in terms of levels of embedded intentionality has found its way into the work of other scholars across both the humanities and sciences. To mention just a few examples: Vermeule integrates it in her argument on "why we care about literary characters" (2010: 62-71) and Palmer builds on it in his 2012 analysis of "storyworlds and groups" (181-186). Carney, Wlodarski, and Dunbar (2014) report experimental evidence partly supporting Zunshine's claims, but add that according to their findings readers seem to value higher levels of embedded intentionality more in "familiar" contexts, such as love stories, compared to

[50] Zunshine emphasises that high sociocognitive complexity does not necessarily involve multiple characters, but can easily occur within the realm of just one character's thoughts, for which she gives the example of Robinson Crusoe imagining what God would think of him thinking about…etcetera (2012: 13). She seems to leave aside that "God" could be seen as another character here. Of course, "God" in this passage only seems to "exist" within the realm of Crusoe's mind, so on the level of the story's here-and-now there might indeed be only one character. However, it should be noted that it is a crucial aspect of embedded mindstates in any context that they are projected *within the scope* of the first one (see also Chapter 1, Section 1.2). So if we are told that A intends that B thinks C wants to get married, and nothing else, all we know is something about A's mindstate, regardless of whether B and C are present in the here-and-now of the story—after all, we are told neither what B thinks nor what C wants, we only know what A *thinks* that B thinks that C wants. This may seem a straightforward issue, but as I will argue in Section 3.3 below, this issue of scope is actually one of the problematic aspects of Zunshine's approach.

contexts where they need to process lots of specific, "unfamiliar" information, as in for example espionage stories.

In what follows I will offer an alternative analysis of the passage of *Mrs Dalloway* targeted by Zunshine. The concepts from narratological theory and cognitive linguistics on which I build will first be introduced briefly in the next section.

3.2.3 The broader picture: coordination of embedded viewpoints

Both linguists and literary scholars have taken an interest in the issue of how mindstates, which they often refer to as "viewpoints" or "perspectives", can be represented in discourse. Literary scholars, mostly within the fields of narratology and (its more recently emerged branch) cognitive poetics, make use of typologies of several forms of "speech and thought representation" (STR), distinguishing for example between direct discourse, indirect discourse, and free indirect discourse (see e.g. Dancygier, 2012; Vandelanotte, 2009; Bal, 2009; Fludernik, 1993). Linguists have rather been concerned with studying the grammatical and semantic means used to realise such forms of STR (see e.g. Dancygier & Sweetser, 2012; Evans, 2010: ch. 4; Tomasello, 2008: ch. 6; Verhagen, 2005).

In recent years, attention has been drawn to structurally different roles played by different linguistic instruments for coordinating viewpoints, across languages and modes of communication as well as within languages and modes. In English, as well as in most other Middle- and Western-European languages, sentence embedding is one of the common ways of attributing propositional content to a subject's point of view.[51] In prototypical cases, a verb of cognition or communication (boldface below) is specified by an embedded clause, for example:

(1) Mrs. Dalloway **said** she would buy the flowers herself.
(*Mrs Dalloway*, Woolf, 1925: 3)

[51] In Chapter 4 I distinguish complementation constructions from several other (though related) patterns that involve sentence embedding. Here, however, I will simply refer to the broad category as "embedded sentences".

(2) She **imagined** the party would be crowded.
(*Mrs Dalloway*, Woolf, 1925: 47)

(3) I **think** I can **imagine** how stunned you must have been.
(Colorless Tsukuru Tazaki and his years of pilgrimage, Murakami, 2014: 51)

As can be observed in ((3), an important feature of the grammatical operation of sentence embedding is that it can be performed recursively. Using this form of viewpoint coordination one can thus stack up large numbers of viewpoints in a single sentence, as is done in Zunshine's paraphrase cited in 3.1 above: "Woolf *intends us to recognize* [...] that Richard *is aware* that Hugh *wants* Lady Bruton and Richard to *think* [...etc.]". However, it is important to note that sentence embedding is only one of many ways of coordinating multiple viewpoints in language, and that it creates a very one-dimensional relationship between these mindstates that easily gets opaque. This relationship will be the topic of Sections 3.3.1 and 3.3.4 below.

In the practice of representing multiple mindstates in discourse, sentence embedding is generally used in combination with an array of other linguistic elements capable of viewpoint coordination (see also Chapters 4 and 5 below). Sweetser (2012: 4-6) provides a classification that includes deictic expressions of place and time, usage of determiners, pronouns and address forms, connectives and evidential markers, presuppositions, and markers of emotion and affection. These elements all differ in how they prompt readers to construe the relationships between the represented mindstates. In the next sections and throughout Chapter 4, I will demonstrate how a polyphonic thoughtscape of interlinked (but not necessarily embedded) mindstates is in the actual text represented using a mix of different elements of viewpoint coordination—or reversely: how the text uses a clever mix of such elements to prompt readers to imagine a polyphonic thoughtscape, rather than a string of embedded layers.

3.2.4 A first example: embedding of perspectives in a novel

Complex mindreading situations tend to emerge gradually in the course of a story rather than being packed into one sentence. Narrative discourse is characterised by, what I have in the previous chapter termed, "expository strategies" capable of conveying highly complicated "thoughtscapes" in a natural and manageable way. An example from Murakami's recent novel *Colorless Tsukuru Tazaki and his years of pilgrimage* (2014) illustrates this nicely, and arguably even provides meta-fictional reflection on the paradoxical nature of the construction of viewpoints in novels: such viewpoints can in some sense be related in very complicated ways, whereas their construction proceeds almost unnoticed most of the time. In Murakami's novel, the main character Tsukuru and his friend Haida regularly have long conversations in the evenings. One night, Haida starts telling a story that his father always used to tell. Throughout the novel, narrating is done in retrospect by a narrator who provides insight in Tsukuru's inner life. The events of the novel are thus seen through Tsukuru's eyes—or in other words: Tsukuru acts as the primary focaliser. Normally the narrator does not directly access mind-states of other characters, but lets the reader construe them where necessary through what Tsukuru thinks or what he hears that others say, which is indicated by quotation marks. However, the quite long story Haida tells about his father is first related through a few sentences between quotation marks, but then continues in unquoted form:

> (4) 'When my father was young, he spent a year wandering around Japan,' Haida began. 'This was at the end of the 1960s [...] I guess people need that sort of stage in their lives.'
>
> That winter Haida's father worked as general handyman at a small hot springs resort in Southern Japan. He really liked the place and decided to stay for a while.
>
> (60-61; quotation marks and tab space in original)

The structure of the story at this point can in some sense be analysed as follows: the narrator provides insight in what Tsukuru hears his friend Haida say.

Haida, in turn, tells what his father said in the past. Haida's father, at that time, was telling what he experienced during an episode earlier in his life, which comprises dialogues occurring at that time and memories of yet earlier times. All in all, five to six layers are in some way active at the same time, while the text simply presents first- and second-order thoughts and feelings of the form "He really liked the place and decided to stay for a while".

The embedding of all these viewpoint layers has gradually taken place in the course of the story's unfolding. Even though it is possible to pinpoint the emerged scaffolding in an analysis, to the average reader only a few details about the entire configuration are relevant to make sense of the presented events. In principle, a cue prompting these details every now and then is enough to keep the configuration sufficiently active in the reader's minds to let the story move on (e.g. "Another person with a color, Tsukuru thought, but said nothing and listened to the rest of the story", 62). However, the layers can still get mixed up despite such cues, as Murakami lets his readers realise when he at some point no longer writes "Haida's father", but "Haida", at places where he is clearly referring to the character of Haida's father in the story Haida is telling to Tsukuru. Just when attentive readers start asking themselves whether this is an inelegant mistake or perhaps something they should try to interpret, the narratological bushfire is smothered:

> (5) Haida stopped and glanced at the clock on the wall. [...] He was, of course, Haida the son, but Haida the father had been his same age in this story, and so the two of them began to overlap in Tsukuru's mind. It was an odd sensation, as if the two distinct temporalities had blended into one. (65)

Murakami clearly takes his place next to Woolf in the category of writers who are at the same time acute literary scholars: he first makes his readers go through the same "odd sensation" that Tsukuru undergoes, thereby blending all the different layers even further, and then he inserts a remark that prompts reflection on what has just happened on a narratological level, even using the technical term "blending" (see Section 3.3.3 below).

The bottom-line of the brief analysis of this first narrative example is that layers of embedding can stack up easily and sometimes go almost unnoticed in

the course of the unfolding of a novel's plot, and that they are usually cued by a mix of quite different linguistic and narrative elements in the text. In order to follow the story, readers have to keep some details in mind of the layered configuration that has emerged, but at least in this case they do not seem to be required to go to extreme lengths in terms of cognitive effort. All they need to know here comes down to questions of the type: "Are we speaking about Haida-the-father or Haida-the-son here?", and would be quite misrepresented by questions of the form: "Does Murakami *intend* that we *think* that Haida-the-son *intends* Tsukuru to *understand* that Haida-the-father *thinks*...(etc.)?" It seems to follow that embedding of viewpoints *per se* in this case is not the factor responsible for complexity or, for that matter, warranting literary quality: it would hardly be problematic, nor would it be a guaranteed literary master move, if the story about Haida-the-father would gradually develop into another story told by *his* father (who would be Haida-the-grandfather), even though the number of embedded layers at that point would theoretically reach seven or eight.

In the next section I will analyse the passage from *Mrs Dalloway*, and propose a different conceptualisation of the complexity involved in representing multiple mindstates in the discourse of this novel.

3.3 Viewpoint layers in *Mrs Dalloway*

In this section I will contrast Zunshine's paraphrase (as cited earlier in the Introduction) to the corresponding excerpt from Woolf's novel itself:

(6)

> Hugh produced his fountain pen; his silver fountain pen, which had done twenty years' service, he said, unscrewing the cap. It was still in perfect order; he had shown it to the makers; there was no reason, they said, why it should ever wear out; which was somehow to Hugh's credit, and to the credit of the sentiments which his pen expressed (**so Richard Dalloway felt**) as Hugh began carefully writing capital letters with rings round them in the margin, and thus marvellously reduced

> Lady Bruton's tangles to sense, to grammar such as the editor of the *Times*, **Lady Bruton felt, watching** the marvellous transformation, must respect.
> (Woolf, 1925; boldface added)

(7)

> (i) Woolf *intends*
> (ii) us to recognize [...]
> (iii) that Richard *is aware*
> (iv) that Hugh *wants*
> (v) Lady Bruton and Richard to *think*
> (vi) that because the makers of the pen *believe*
> that it will never wear out
> (vii) the editors of the *Times* will *respect and*
> *publish* the ideas recorded by this pen.
> (Zunshine, 2012: 206-207; italics in original, numbering and indents added)

Whereas the narrative in ((6) is perfectly readable, also when seeing it for the first time, the paraphrase in ((7) is highly opaque. As argued in Chapter 2, sentence grammar is well capable of handling up to two or three orders of intentionality, but if more orders are involved, "narrative takes over". Here, however, I am interested in the lower-level linguistic and narratological phenomena that realise viewpoint construction and management, rather than in these broader expository strategies. To that effect, I will compare the narrative passage cited in ((6) and the paraphrase cited in ((7), looking at differences in their construction of (what is supposed to be, in some respect) the same situation by applying Dancygier's (2012) framework of "narrative spaces". Choices made on the level of these linguistic and narratological phenomena influence the way in which several viewpoint layers become mutually embedded and interlinked. As will turn out, this affects not only meaning, but also the ease (or difficulty) with which the emerging network of perspectives can be processed.

3.3.1 The first three layers

The paraphrase, which Zunshine forms to analyse the passage from the novel in terms of orders of intentionality, comprises seven viewpoint layers.[52] The first two concern the relationship between the author and the reader: they refer to the basic communicative situation of reading a novel. However, rather than being coordinated by linguistic cues *in* the text, these viewpoint layers come implicitly with the genre. Although readers who open a novel may in a very abstract sense be "aware" that the text written on the pages contains a story that the author *intends* them to *believe*, in the actual practice of reading these layers are not as much activated as is suggested by Zunshine's paraphrase. For a reader (or a literary critic setting aside for a moment all axioms that come with the "death of the author"; Barthes, 1967), it is at any point in the text possible to stop reading and consider explicitly what the author may have intended one to believe by writing, for example, that someone carries a bunch of flowers "like a weapon" (*Mrs Dalloway*: 102). But this does not mean that the minds of readers are burdened with the cognitive load of this task during the average reading process. Thanks to their experience with the basic communicative situation of the novel, they can, as it were, start processing from beyond the levels (i) and (ii) in ((7) (see also Dancygier, 2012a, for the idea of the basic communicative situation as a *frame*).

Layer three of ((7) brings one of the novel's characters into play: Richard Dalloway. Abstracting from the positions of the author and reader, Zunshine thus reconstructs the scene from his point of view: after all, the nature of embedding clauses as is done in (7) is such that every next clause falls under the scope of the former. To some extent, the text does indeed give rise to this choice: the remark between parentheses "so Richard Dalloway felt" (boldface in (6)) invites readers to see Hugh writing the letter from Richard's perspective. Or, in narratological terms: it is true that Richard acts as the focaliser of a part of the scene. However, note that in the original passage, this is only the case for a short moment. Focalisation was with Lady Bruton before, and is passed back

[52] Zunshine herself suggests that this paraphrase is sixth-level intentionality; her italics suggest that she takes (i) and (ii) as one level. However, this is not consistent with counting in other paraphrases she provides (2012: 207-207). Since "*intends us to recognize*" contains an extra embedding, I decided to count this as two orders, which brings us at a total of seven.

to her immediately after Richard has briefly taken the floor, witness the clause "Lady Bruton felt, watching the marvellous transformation" in (6).

This is an important observation: whereas the nature of sentence embedding makes all subsequent layers dependent on Richard's perspective, in the original novel's text, as perspectives shift, these layers become interlinked in various ways *without* necessarily being embedded into one another.

3.3.2 Shifting perspectives

In *Mrs Dalloway* an omniscient narrator thus provides a wealth of insight into the inner lives of various characters, resulting in a "360-degree view" of the novel's thoughtscape. In order to fully understand and, for that matter, enjoy the passage cited in ((6), in the novel on page 96, some knowledge of the previous 95 pages is needed. After all, lots of details about the characters, including their sympathies and antipathies towards one another, will by this time have been established. For example, if we limit ourselves to the excerpt as cited in (6) only, it appears that narration starts from a "neutral" position, as if there were a camera recording from above:

(8) Hugh produced his fountain pen

Indeed, in narratological terms one would say that it is the omniscient narrator "witnessing" Hugh's action and reporting it to the reader. However, for readers who have just read what happened before this point, everything Hugh does and says here is already viewed in a specific way by Lady Bruton and Richard Dalloway. Earlier that afternoon, Richard and Hugh had arrived for lunch at Lady Bruton's house, where they were in some sense invited on false pretences: in fact Lady Bruton wanted their help with the writing of a letter to the *Times* to publicly express her thoughts on a political matter. Before lunch is eventually served, readers already share in Lady Bruton's thoughts about Hugh and Richard, including her opinion that Hugh "had been remarkably kind" (yet she forgot on which occasion) but that she preferred Richard, who was "made of much finer material" (91). This information, combined with knowledge that the readers have acquired before of who these characters are and how they think

about each other, clearly invites the attribution of more sophisticated thoughts than can be explained from analysing only the textual cues within the boundaries of (7). In the current and following sections I will nevertheless look at the cited passage in relative isolation, concentrating on the linguistic phenomena coordinating the different viewpoints involved in the depicted scene. Eventually, in Section 3.4, a wider view on the text will be taken into account.

After the narrator has told us how Hugh produced his pen, a form of indirect discourse is used to relate a comment on this pen made by Hugh.[53] Linguistic cues signalling viewpoint management (in narrative space theory referred to as *space builders*, see below) are printed in boldface:

(9) his silver fountain pen, which had done twenty years' service, **he said**, unscrewing the cap
(boldface added here and in 10-12 below)

It should be noted how the transition from the narrator's voice to Hugh's voice and back is ingeniously cued here. First it is clear that the narrator is speaking, describing Hugh's action from a third-person perspective: "Hugh produced his fountain pen". The sentence then continues with a further specification of this pen, "his silver fountain pen, which had done twenty years' service", which could in principle still be a comment added by the narrator. However, when the inquit-formula "he said" follows next, it becomes clear that the narrator *only now* takes over again. As a consequence, readers have to do a "double take" on what they have just read, realising that it must have been Hugh saying that his silver fountain pen had done twenty years' service while unscrewing the cap (in fact, the reader may well realise by the end of the sentence that the addition of "silver" should already have been a cue of the perspective of the slightly pompous Hugh). In a way, this is a "special effect" highlighting the wandering

[53] More precisely, (9) is an example of what I will term an "inquit-construction" in Chapter 4, named after the inquit-formula (in this case "he said") that attributes the propositional content (in this case "twenty years' service" or "silver fountain pen, which had done twenty years' service", depending on interpretation) to the perspective of a particular discourse participant (in this case Hugh).

nature of narration here: the need of doing a double take may remind readers that perspectives are constantly shifting.[54]

Next, the text continues as free indirect discourse, within which Hugh reports, in turn using indirect discourse, what the makers of the pen have said:

> (10) It was still in perfect order; **he had shown** it to the makers; there was no reason, **they said,** why it should ever wear out;

From here perspective seems to shift back to the narrator, but in the same way as with Hugh's voice in (9), readers have to do a double take as soon as they reach the comment between parentheses, which makes it clear that they had been reading about Richard's thoughts:

> (11) which was somehow to Hugh's credit, and to the credit of the sentiments which his pen expressed (**so Richard Dalloway felt**) as Hugh began carefully writing capital letters with rings round them in the margin, and thus marvellously reduced Lady Bruton's tangles to sense,

Finally, this is seamlessly followed by the shift towards Lady Bruton's viewpoint:

> (12) to grammar such as the editor of the *Times*, **Lady Bruton felt**, **watching** the marvellous transformation, **must respect**.

To summarise, the excerpt from the text in (6) in fact describes three "scenes":

> (a) the present in Lady Bruton's house: Hugh composing a letter to the *New York Times*, based on a series of opinions (the "tangles") produced by Lady Bruton;

[54] See in this context also Coulson's discussion of "frame shifting" (2001). Moreover, note that in narratology the phenomenon of blending different levels of narration is known as metalepsis (De Jong, 2009; Genette 1972). In its classic form, it involves a narrator becoming a participant in the world of the characters, or one of the characters interfering with the narrator's business of telling the story. The "double take" that readers are cued to do in (9) is arguably a clever variation on this form.

(b) the past of the fountain pen: it having been in service for twenty years and Hugh showing it to the makers; and

(c) the imagined future: the editors of the *Times* receiving and reading the letter.

Various means of viewpoint coordination are used to let the readers "experience" aspects of these scenes from different perspectives: indirect discourse and free indirect discourse serve to represent speech by Hugh and by the makers of the pen; indirect thought, supported by inquit-formulas using the verb "to feel", attribute observations to the viewpoints of Richard Dalloway and Lady Bruton; and indirect thought is used to include the presumed opinion of the *Times* editors. As will be detailed in the next section: rather than evoking a series of embedded perspectives viewed from one single vantage point, the text prompts readers to imagine what I have termed a thoughtscape, a network of different perspectives that are interlinked and partly embedded in a variety of ways.

3.3.3 Narrative spaces and blending

The three scenes distinguished at the end of the previous section differ in their setting in space and time, participants, and the viewpoints from which they are perceived and narrated. The writing of the letter (a) is the "actual" setting at that point in the novel's plot, within which the other two scenes are evoked by character's words and thoughts. The twenty years' service of the pen and the event of showing it to its makers (b) are referred to by Hugh. The editors receiving and reading the letter (c) can be seen as a hypothesised future event within Lady Bruton's thoughts.

The configuration of these scenes as prompted by the text must somehow imply the embedding of mental states: after all, there are characters (Hugh; Richard; Lady Bruton) thinking of persons (the makers of the pen; the editors of the *Times*) having a particular thought or opinion (that the pen will not wear out; that the letter is respectable). However, focussing on embedding *per se* does not provide an adequate picture of how all the mindstates involved in the passage cited in (6) are mutually related and interlinked. For a detailed analysis supporting this claim I will rely on Dancygier's *narrative-spaces framework* (2012;

2008). This framework offers a version of mental space theory (Fauconnier, 1997; 1985) tailored for narrative texts. Like a mental space, a narrative space is a hypothetical subdivision in a language user's mental activities, prompted by linguistic expressions, and used in the process of meaning construction. However, whereas mental spaces typically form "ad hoc"-structures that are constantly modified or replaced in the course of interaction, configurations of narrative spaces can persist throughout an entire story once they have been prompted. Each narrative space is characterised by a particular set of features such as time, space, cultural norms, language spoken, or participants involved in either narration (narrators/focalisers), action (characters), or both (Dancygier, 2012: 35-37). These features are open to further elaboration by all kinds of local linguistic choices, such as use of sentence embedding, grammatical features such as tense or modality, use of pronouns, typographical cues (e.g. quotation marks), etcetera.

For an important part, construction and elaboration of narrative spaces has been argued to take place through *blending* (Turner and Fauconnier, 1995; Coulson, 2001). The general idea of blending is that two inputs with an established conceptual structure and content are integrated into an emergent *blended space* or *blend*. The blend has properties inherited from the inputs, as well as new structure and content of its own. Once a narrative space has been prompted and possibly further structured and enriched, it can as a whole be blended with another narrative space. The result of all the construction and blending processes is what Dancygier refers to as the *emerging story*: this is what the reader understands after having read and processed the text up until a particular point. The emerging story is thus a "moving end result".

As an example, consider this excerpt from the very beginning of Murakami's novel *Norwegian Wood*:

(13)

> I was 37 then, strapped in my seat as the huge 747 plunged through dense cloud cover on approach to Hamburg airport.
>
> [...] Once the plane was on the ground, soft music began to flow from the ceiling speakers: a sweet orchestral cover version of the Beatles' "Norwegian Wood". The melody [...] hit me harder than ever.

> I bent forward, my face in my hands to keep my skull from splitting open. Before long one of the German stewardesses approached and asked in English whether I were sick.
>
> "No," I said, "just dizzy."
>
> [...] She smiled and left.
>
> [...] The plane reached the gate. People began unfastening their seatbelts and pulling luggage from the overhead lockers, and all the while I was in the meadow. I could smell the grass, feel the wind on my face, hear the cries of the birds. Autumn 1969, and soon I would be 20. (Murakami, *Norwegian Wood*, 2000)

This story excerpt stages an "I"-narrator describing a past event: landing at Hamburg airport when he was 37 years old, and within that event being reminded of a moment even further back, from when he was almost 20. In Dancygier's terms this can be described as a structure with three narrative spaces, where the first space includes the second and the second includes the third. Through the first words ("I was 37 *then*") a narrative space is prompted comprising the "I" at the moment of telling the story. No other details about this first space are mentioned. Subsequently, a second narrative space is prompted within the first one, comprising two participants (the "I" and the stewardess), a particular spatio-temporal setting (some day 18 years after 1969 in a plane landing at Hamburg airport), first-person narration, and past tense. A few sentences later, the episode from further back (a day in a meadow in 1969) is inserted in the form of a memory (in conventional narratological terms this would be referred to as a *flashback* or *analepsis*). This means that another space is prompted within the two existing ones. The result on the level of the emerging story is a blend of the three narrative spaces: while the remembered episode from 1969 comes into focus, the setting of being in the plane at Hamburg airport also persists in some form, as does the here-and-now of the "I" telling the story of when he was 37. The reader understands the text as having a layered structure, where each part is set against the background of the other. Schematically:

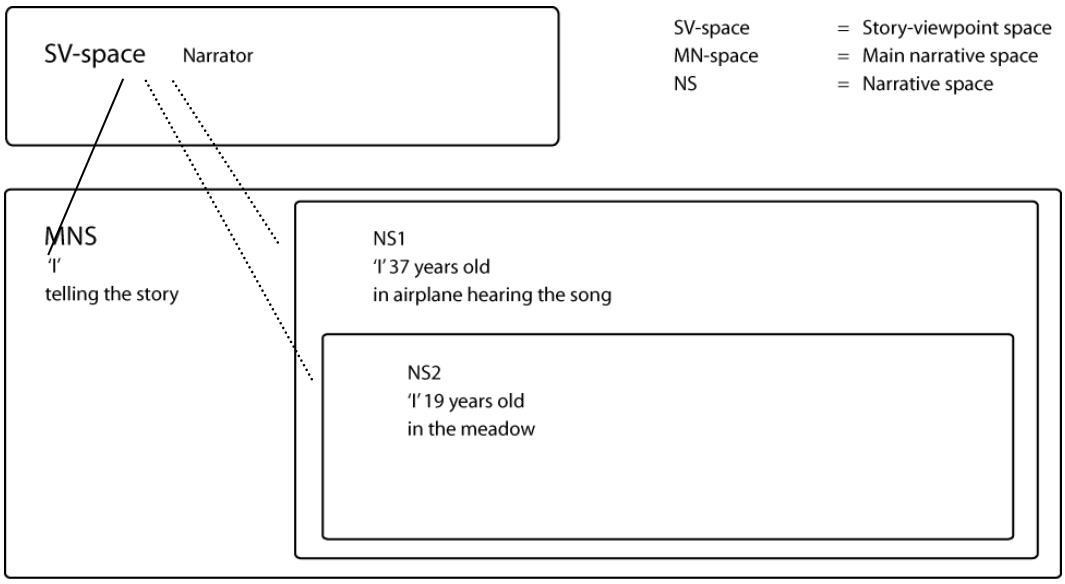


Figure 1 – Schematic depiction of the narrative-spaces configuration prompted by excerpt (13). Following Dancygier, I assume a separate space for the narrator, called the story-viewpoint space (SV-space; 2012: 64-75). It can be seen as an "overarching" narrative space, housing the vantage point that has all the other spaces in its scope. The solid line indicates identity: the narrator in the SV-space is identical to the primary focaliser ("I") in the MN-space. The dashed lines between the narrator in the SV-space and "I-37-years-old" and "I-19-years-old" indicate that focalisation at some point shifts to these perspectives.

Note that usage of the blending framework is especially adequate here, because the story at the end of the excerpt in ((13) has neither just the content and structure of the initial "now" of telling the story, nor just that of the situation in the plane or that of the day in the meadow, nor is it simply the sum of the three—it has elements of all three spaces, as well as newly emerged properties of its own: "Norwegian Wood" is now no longer just a Beatles song, but it is charged with aspects from the main character's personal history.

This example captures an important feature of stories in general: at any point between their beginning and ending they exhibit a unique emerging story, a particular structure and content unique to this point in the narrative. At the same time, this emerging story results from all the elements that contributed to its construction and development so far. In other words: any point in a story is somehow dependent on the set of preceding points in it. Still, it is clear that the complexity at any particular point does not equal the sum of

all the complexities of the earlier points. Some aspects are introduced and remain a clearly visible part of the emerging story until the ending, others modify the emerging story in some persisting way without remaining visible in their initial form, and yet others affect the emerging story only temporarily or fade away as the story progresses (in blending theory, the process in which some aspects of the input spaces are transferred to the blended space while others are, sometimes temporarily, dropped, is referred to as *compression*—see Dancygier, 2012; this concept will be used more extensively in Chapter 4).

3.3.4 Back to Mrs Dalloway

With this introduction to the narrative-spaces framework in mind it is time to return to *Mrs Dalloway*. Following the logic explained above, the following schema can be drawn:

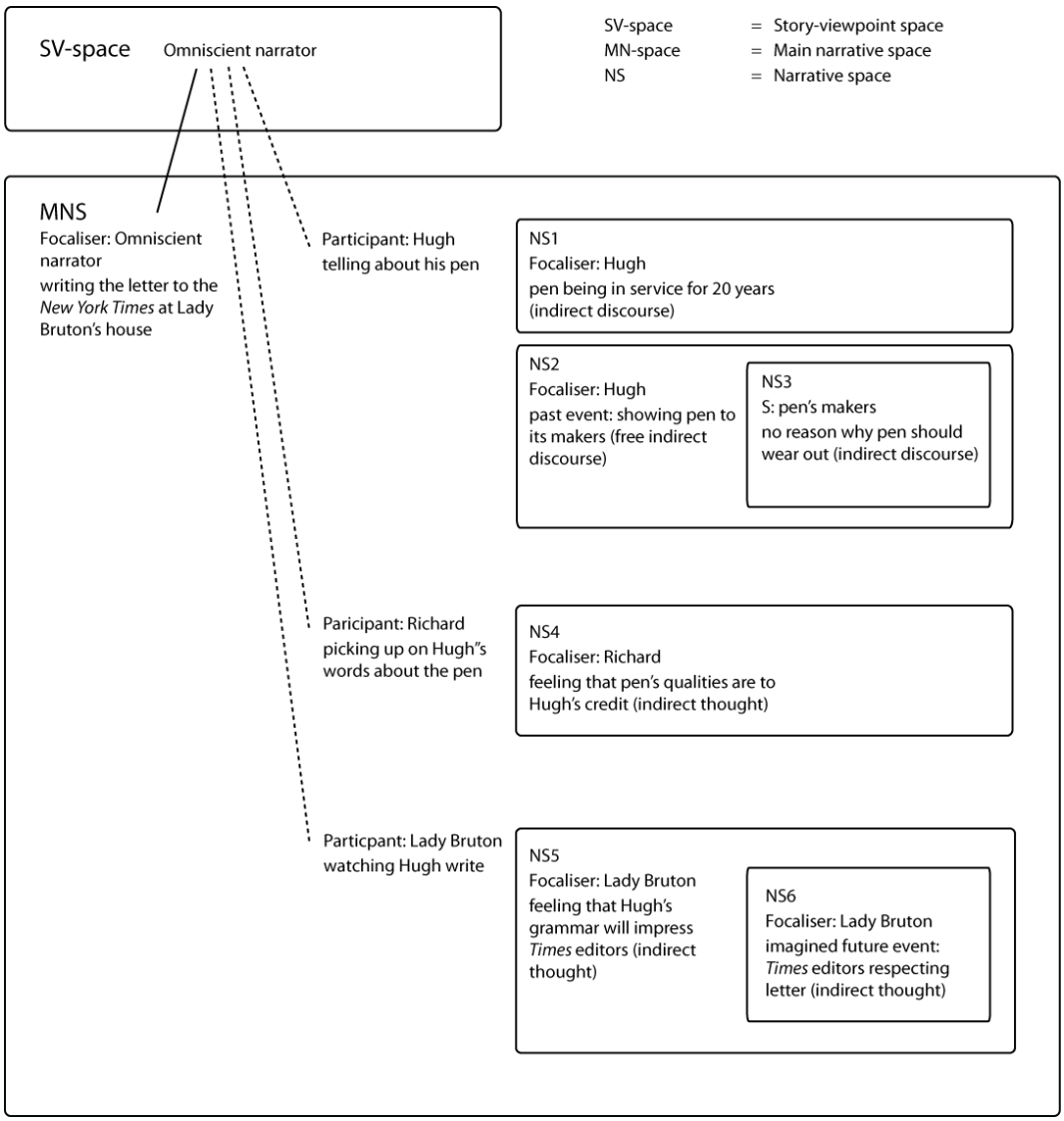


Figure 2 - Schematic depiction of the narrative-spaces configuration prompted by excerpt (6).

As in Figure 1, the solid line indicates identity: the omniscient narrator in the SV-space is identical to the primary focaliser in the MN-space. The dashed lines between the narrator in the SV-space and the characters participating in the MNS indicate that focalisation sometimes shifts to their perspective, with their thoughts and inner lives being accessed by the narrator and presented to the reader. This is what distinguishes them from the makers of the pen and the editors of the *Times*, who are also characters, but hold only thoughts and

opinions that are attributed to them; that is: their inner lives are presented under the scope of the viewpoints of Hugh and Lady Bruton instead of being directly accessed by the narrator.

From this schematic analysis of the Woolf-passage, four conclusions can be drawn. First, it can be seen that linguistic choices made on the micro-level either structure an existing narrative space or prompt a new one. For example, NS1 is prompted by the inquit-formula "he said" (boldface in (9)) and NS4 is prompted by the comment between parenthesis "(so Richard Dallow felt)" (boldface in (10)). Secondly, it can be noticed that the narrative spaces have features (spatio-temporal setting, participating characters, point of view) that are partly different and partly overlapping or identical. Thirdly, they can naturally be grouped in what could be called "intermediate-level blends" or "scenes", which in turn combine into the "highest-level blend" of the emerging story. For example: NS1, NS2, and NS3 together form the scene of Hugh telling about his pen, whereas NS5 and NS6 combine naturally into the scene of Lady Bruton watching Hugh write and consequently imaging what the editors of the *Times* may think. Fourthly, and related to this third point, Figure 2 depicts how some of the spaces are embedded into one another, while others are linked in different ways, and others again are independent of one another. NS1 and NS2 are embedded in the MNS; NS4 is triggered by and provides a perspective on NS1, NS2, and NS3; and NS6 is embedded in NS5, which is embedded in the MNS. Richard's thoughts in NS4 might be triggered by Hugh's words in NS2 and NS3, but they are not part of Hugh's perspective. Also, Lady Bruton's thoughts and expectations in NS5 and NS6 are particular to her and not shared by anyone of the other characters.

The result on the highest level of the emerging story is indeed an increasingly rich "360-degree view" of the situation in Lady Bruton's house, with a "camera" that seems to turn smoothly from one position to another. The readers first find Hugh in a focalising and speaking role, and the other two participants in the background. Next, or perhaps rather simultaneously, they see Richard in a more active role, being the focaliser, watching and evaluating Hugh while writing. Finally they see Lady Bruton take the floor, having thoughts about the *Times* editors' opinion, while the setting of Hugh writing the letter also remains in focus, but Richard's thoughts are dropped. Once again: all

these viewpoints are *interlinked* in several ways, but this does not mean that they are *embedded* into one another.

Whereas Figure 2 depicts the narrative space configuration as prompted by the original passage, Figure 3 below depicts the paraphrase as formed by Zunshine in her analysis of this passage:

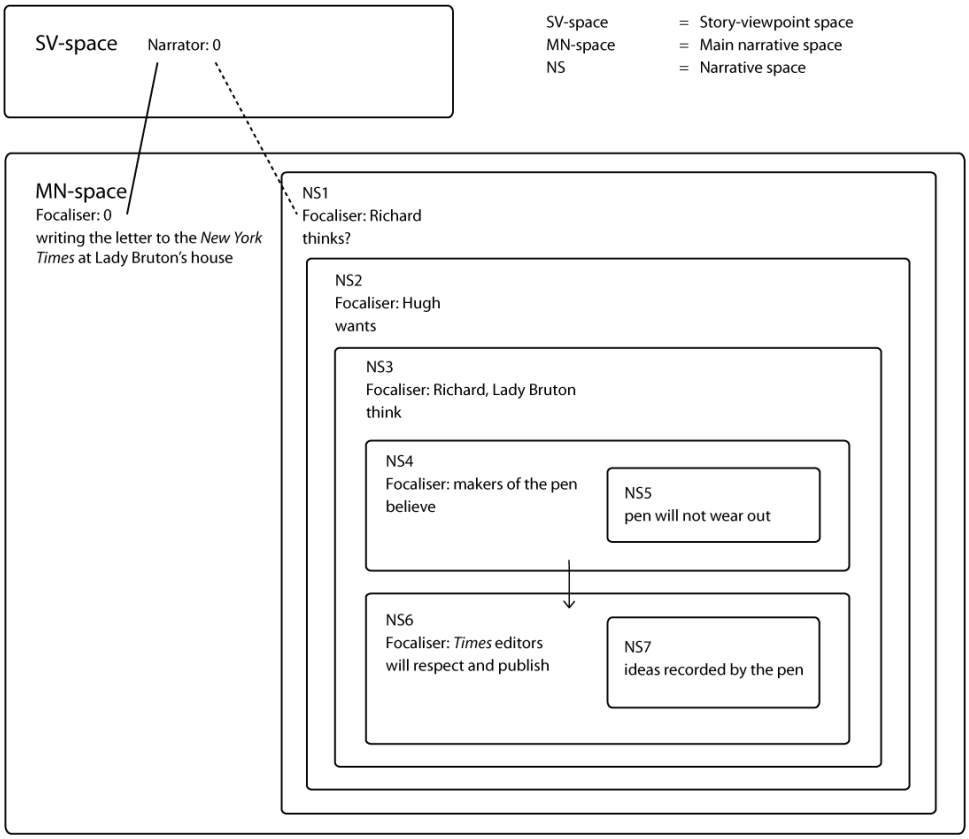


Figure 3 - Schematic depiction of the narrative-spaces configuration prompted by excerpt (7).

Since the narrator nowhere narrates or focalises independently, and is thus completely off-stage, a "0"-label is given in the SV- and MN-spaces (cf. Dancygier, 2012: 66-68). Furthermore, each of the embedded clauses (abstracting from the author and reader) prompt a new space: "...that Richard **is aware** that Hugh **wants** Lady Bruton and Richard **to think** that because the makers of the pen **believe** that it **will never wear out** the editors of the *Times* **will respect and publish** the ideas **recorded** by this pen". The causal link construed through the connective "because" in the paraphrase is indicated in

the figure by an arrow between NS4/NS5 and NS6/NS7. Except for these causally linked spaces, all others are set up *within* the previous space, as a consequence of which all spaces fall under the scope of Richard's viewpoint. This is indicated by the dashed line between him and the SV-space, showing that he is the only agent to which focalisation is "passed" from the narrator. The resulting narrative-spaces configuration in Figure 3 reflects the one-dimensional, opaque "stack" of perspectives as construed by the recursively applied sentence embedding in Zunshine's paraphrase.

3.4 Discussion and concluding remarks

As discussed in Section 3.2, Zunshine suggests that Woolf's *Mrs Dalloway* exhibits a high degree of, what she refers to as, sociocognitive complexity. She starts from what I have labelled as assumption (i) in Section 3.2.2: that this complexity must be conceptualised as a series of embedded layers. Next, she selects a passage in order to "map [it] out in terms of the nested levels of intentionality" (2012: 206). After having discussed a few smaller "irreducible units" of embedded intentionality (such as "Hugh says that the makers of the pen think it will never wear out") she sets out to "move up to those that capture as much of the whole narrative gestalt [of the cited passage] as possible" (2012: 206). This is the context in which she forms the paraphrase cited in (7) above. Clearly, this paraphrase was not intended to replace the original text: she acknowledges that it is only one way of mapping out the targeted passage. However, and that leads to her assumption (ii), she argues that in order to grasp the full meaning of the passage, readers first have to accomplish the highly cognitively demanding task of processing "several sequences that embed at least five levels of intentionality" (2012: 207).

In fact, the paraphrase does cover a quite subtle aspect of what happens in the cited passage of the novel. By mentioning the age and expensive material of his pen and the opinion of its manufacturers when they saw it, Hugh does more than only sharing information with Richard and Lady Bruton. We can safely

ascribe to him the intention of persuading his interlocutors that he is a pro at writing, satisfied only with the best tools to support this profession (or something to that effect). If we do so, we add extra complexity to our understanding of the text while reading: instead of only seeing that Hugh possesses a great pen, we also attribute (what we presume to be) his implied motives for drawing attention to this fact. The text hints that Richard in any case *is* sensitive to this implication, since his feeling that the quality of the pen somehow adds to Hugh's and the letter's credit is put on stage by the lines "which was somehow to Hugh's credit, and to the credit of the sentiments which his pen expressed (so Richard Dalloway felt)". This observation seems to be what Zunshine has woven into the paraphrase: it covers Richard's sensitivity to Hugh's presumed intention to imply his expertise (layer iii-vi in (7)).[55] However, if we stay close to the text, we cannot say whether Lady Bruton has this sensitivity too; we are only informed about *her* positive feelings concerning Hugh's mastery of grammar and *her* expectation that it will gladden the editors of the *Times* too. Also, the text does not mention (or strongly imply) any of Richard's expectations of what the editors will think. This is where my conceptualisation of the text providing a "360-degree view of the thoughtscape" works better than that of the layered structure as suggested by the paraphrase: the latter misrepresents Lady Bruton's idea of the editors' opinion (layer vii in (7)) as following directly from her (and Richard's) understanding of Hugh's presumed intentions behind his remarks about the pen. In the same vein, the text does not mention (or strongly imply) Richard's expectations of what the editors will think, which is thus also misrepresented by the paraphrase. This leads to similar conclusions as drawn earlier: by embedding into one another all the viewpoints held by or ascribed to the characters, the paraphrase does not seem to connect the scenes presented by the text in the right way: it distorts the interlinked though mostly independent nature of the viewpoints.

A perspective on the analysis offered in this chapter that is worth working out in more detail in the future, follows from the work of Bakhtin. One of his

[55] Incidentally, there is still a difference between Richard seeing through Hugh's intention to brag and Richard being affected and truly impressed by Hugh's bragging—deciding between these two nuances is not possible on the basis of the fragment or the paraphrase. This is different, though, for readers of the entire novel, who will know enough about the two characters to see that the first of the two options is the more likely...

central points is that characters and their perspectives in a novel should be seen as "dialogising voices", interacting also beyond what is literally written down in the text (see e.g. Mey, 1999: 153-154 and his references to Bakhtin, 1992). They "populate" a text, and even though an author or narrator can give her best attempt to "orchestrate" their voices, she will never be able to fully control them (cf. the concept of "heteroglossia"; Bakhtin, 1992). One important reason for this, according to Bakhtin and his adherents, is that readers will always, consciously or unconsciously, attribute all kinds of thoughts and intentions to characters. It is for sure thinkable that a text can bring a reader in a position where it is both hard to access a particular character's thoughts and necessary to do so in order to be able to understand the story—in that case, it is also thinkable that reading would be cognitively demanding. However, as far as the literary excerpts analysed in this chapter are concerned, quite the opposite situation seems to obtain: texts are full of elements that make the contents of different mindstates and their mutual relations *accessible* instead of opaque, and, thinking in the way proposed by Bakhtin, readers seem prone to (over)attributing intentional relationships when interacting with the text easily rather than by virtue of hard cognitive effort.

This finally takes us back to assumption (iii) defined in Section 3.2.2 above: that readers' appreciation of a literary work is affected by the amount of mindreading complexity it contains. Rather than being drawn to fictional texts because it is so hard to figure out what all the characters are thinking that this pushes us to the limits of our abilities, we might just as well conclude that fiction attract us because it takes us relatively *little* effort to get access to rich representation of others' inner lives, compared to real-life settings where we usually do not get such "360-degree spectator sight". Like in our daily social environments, in a novel such as *Mrs Dalloway* we are confronted with a rich thoughtscape, constituted by a polyphony of different voices, that underlies all events and interactions. However, unlike in our daily lives, in *Mrs Dalloway* the text provides us with a full panoramic tour of all the mental space we normally don't get to access so easily. Put differently: as readers of fiction such as Woolf's, we can sit back in our chairs lazily and enjoy the thoughtscape, while the narrator does most of the mindreading for us.

Chapter 4

Viewpoint Packages: linguistic tools for communicating and processing complex "thoughtscapes"

A famous athlete shoots dead his girlfriend at night in his house. The next morning, readers of newspapers all over the world find out that a spokesperson *states* that police officials *declare* that the athlete *claims* that he *thought* that he was shooting at a burglar, while the responsible police detective *claims* that he *knew* it was his girlfriend. Clearly, this complex "thoughtscape" is not represented in the news media in this way, using multiply-embedded sentences. I demonstrate that the representation of the involved mindstates relies substantially on lexical units implying viewpoint layers, such as *allegedly*, *accidentally*, *mistake*, and *to confirm*. I introduce and discuss the concept "viewpoint package", building on an existing framework that deals with meaning construction more generally (Dancygier's narrative spaces) and on one relevant account focussing on a particular part of speech (Vandelanotte's framing adjectives). Viewpoint packages allow for efficient coordination of multiple interrelated viewpoints in discourse, while regulating (audience's perception of) the commitment various discourse participants make to parts of the presented contents. I end by tentatively suggesting that viewpoint packages qualify as "tools for thinking": knowing their use in language may serve not only communication, but also support cognitive processing of complex thoughtscapes.

4.1 Introduction

On Valentine's Day 2013 a dramatic event made headlines all over the world: early that morning, a famous Olympic athlete shot dead his girlfriend through the bathroom door. The question that immediately perturbed everyone was whether it was murder, or a fatal accident: the athlete claimed that he had mistaken her for a burglar, while the police arrested him on the charge of having killed her wittingly and on purpose.[56] The crucial point of debate in this case was thus not whether he shot her or not, but whether he knew he would be doing so when he pulled the trigger.

This means that the choice between accident or murder coincides completely with the construal of the athlete's mindstate at a particular moment during the night of the shooting. Ultimately, this construal was made by a judge in court, based on information from forensic research, interrogations, witness reports, and so on. But besides that, and from the very first day after the incident, thousands of people have made such construals for themselves, mostly relying on cues presented by the news media. Although the main focus is clearly on what Pistorius thought, knew, and intended during the night of the shooting, most of these cues do not directly concern the athlete's mindstates. Rather, they add up to a complex "thoughtscape", a network of mutually embedded and interlinked viewpoints that are in some way relevant to the case: news media *suggest* that various sources *report* that the athlete *claimed* that he *thought* that his girlfriend was still in bed and not behind the bathroom door when he fired his gun. At the same time they *report* that a spokesperson *declared* that police officials *considered* it to be likely (or at least to a high degree possible) that the athlete *did know* that his girlfriend was behind the bathroom door, and that he thus *intended* to kill her. The news media also *report* what witnesses *claim* to have heard or seen, or what various sources *report* that family members *have declared*.

[56] Background to the case: South-African athlete Oscar Pistorius shot and killed his girlfriend Reeva Steenkamp on February 14th, 2013. Pistorius is a sports icon also known as "the Blade Runner"; his legs were amputated and yet he became a sprinter using carbon-blade prosthetic legs. In the aftermath of the killing, news media have frequently reported details of the court case, police investigations, personal life of Pistorius and Steenkamp, etcetera.

This entire thoughtscape is somehow represented in the headlines and articles about the case. Partly this is done using means of viewpoint coordination that are well-accounted for in the literature on speech and thought representation, such as indirect speech or free indirect speech. Consider, for example, the following passage from a press release on the morning after the shooting:

(1) Athlete Oscar Pistorius allegedly accidentally shot dead his girlfriend at his house in Pretoria on Thursday morning, Beeld.com reported.
(SAPA, 'Oscar Pistorius shoots girlfriend: report', 14 February 2013)[57]

Using a form of indirect speech, the propositional content of the reported clause "Athlete Oscar Pistorius…Thursday morning" is attributed to the perspective of *Beeld.com*. Yet there are more viewpoints coordinated in this sentence: through the adverbs "allegedly" and "accidentally", readers of (1) also learn that a source other than *Beeld.com* claimed that Pistorius killed his girlfriend without intending to. The information that Pistorius shot his girlfriend is clearly the critical bit, but it comes embedded in a complex of viewpoints that could be depicted as follows:

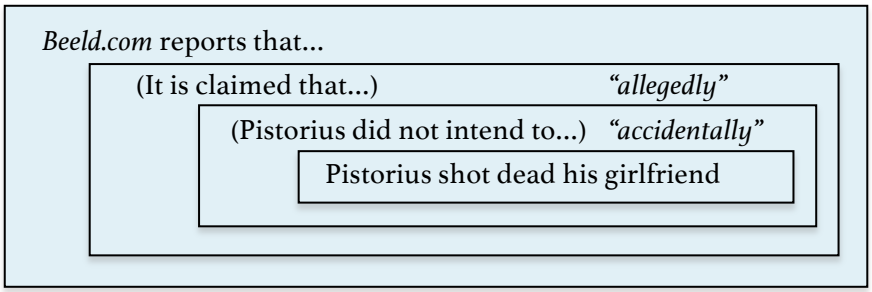


Figure 1

[57] The reports published on 14 February 2013 were retrieved on 21 February 2013 using Lexis Nexis (http://academic.lexisnexis.nl).

In this chapter I will contrast the two ways in which viewpoint layers are introduced into discourse as exemplified here: through a form of indirect discourse on the one hand, and through the use of single words (or lexical units) implying one or more viewpoint layers, such as *alleged(ly)*, *accident(ally)*, *mistake(n)*, *to confirm*, on the other hand. Various aspects of the working of such words have been studied in Vandelanotte's research on "framing adjectives" (2002; 2007) and in approaches dealing with meaning construction more generally (framing, blending, mental space theory).[58] Fauconnier (1997) makes the case that newspaper articles most of the time rely heavily on background knowledge pre-existent in readers. Given that this knowledge is for a large part structured in frames, journalists can activate a whole network of relevant information in readers by mentioning only a few well-chosen cues. This information is then available for interpreting the news details given in the article. The processes of meaning construction focussed on in this chapter could be considered a viewpoint-specific subcategory of framing in this sense. I propose to distinguish a set of words, across part-of-speech boundaries, capable of what I will refer to as "holistic" viewpoint coordination. The label that I propose for this set is *viewpoint packages*. Viewpoint packages, such as *alleged(ly)* and *accident(ally)*, activate readers' background knowledge related to viewpoint, which can then be assimilated with contextual details.

In addition, in my final section I will address the relation between linguistic discourse and cognitive processing by suggesting that viewpoint packages can be seen as instantiations of what Dennett calls "tools for thinking" (2000; 2013): knowing their use in language may alleviate the burden of handling multiple perspectives cognitively.

[58] It should be noted that "framing" in "framing adjective" goes back to a different tradition of using this term: see McGregor (1997: esp. 66-67).

4.2 Viewpoint, mindstates, and the human ability to read minds

"Viewpoint" forms an object of study situated on a busy crossroads of disciplines. Whereas concepts such as "mindstate" and "intentionality" go back a long way in philosophy of mind, "viewpoint" and "perspective" have been traditional topics in the study of language and literature. During the past decades, psychologists and cognitive scientists have taken up the subject, and in recent years there has been an increasing number of attempts to bring insights from different angles together. The approach proposed in this chapter can be seen as an attempt to pursue this line.

In what follows, by a "mindstate" or "intentional state" I mean a particular belief, intention, desire (etc.) held by an animate individual with respect to a state of affairs in the outside world (much in the fashion of Dennett's *intentionality*; 1987). I use "viewpoint" and (in this chapter synonymously) "perspective" as indicating the broader, overall take a person, group, or institution (e.g. a newspaper) has on a certain part of that world. Intentional states or mindstates are thus, as it were, atomic "snapshots" of a subject's relation to an object; a perspective or viewpoint comprises the broader total of an actor's subjectivity of which intentional states are isolated parts.

By definition, neither mindstates nor viewpoints/perspectives held by others can be accessed directly, but they can be appraised through an inferential process based on behavioural cues (including linguistic utterances) and immediate circumstances. This latter process is often referred to as "theory of mind" or "mindreading" (for an overview see Apperly, 2011). Several cognitively-oriented linguists and literary scholars have pointed out that the process by which language users form an understanding of the viewpoints and mindstates of people or characters referred to in discourse, can be seen as a special case of theory of mind or mindreading.[59] In the physical presence of an interlocutor (say, when speaking with John), we have direct access to verbal and

[59] See e.g. Verhagen (2005). For fiction see Palmer (2004); Budelmann and Easterling (2010); Dancygier (2012); Sluiter, Corthals, Van Duijn, and Verheij (2013); Cefalu (2014); see also Chapter 2 and 3.

non-verbal cues that may guide our inferential process regarding John's mindstates. However, when John is not physically present but referred to by Mary in a conversation we have with her (or a story narrated by her), a similar inferential process can be triggered through cues in Mary's speech that enable us to construe John's mindstates or viewpoint. Applied to our object of study in this chapter, understanding a thoughtscape on the basis of news reports is thus seen as a form of mindreading in which textual cues guide the inferential processes within readers.

In the next section I will investigate this process more closely, while bringing elements from a variety of frameworks together. As a first step in my analysis I will distinguish between two ways of introducing viewpoint complexity into discourse: *compositionally* versus *holistically*.

4.3 Constructing complexity*

4.3.1 Compositional complexity: a literary example

Regarding relevant approaches to meaning construction, in particular mental-space theory, framing, and blending, we will mostly rely on Dancygier's narrative-centred synthesis offered in her 2012 book *The Language of Stories*. Her narrative-spaces framework forms a useful tool for analysing how texts can represent a complex thoughtscape. The central issue in this framework is how cues at the micro-level of a text ultimately yield an increasingly rich and complexly structured story at the macro-level. Dancygier builds on mental space theory and offers a related though newly devised core concept: the *narrative space* (see Dancygier, 2012: esp. chapter 2; for mental space theory see Fauconnier, 1997; 1985). Like a mental space, a narrative space is a hypothetical subdivision in a language user's mental activities, prompted by linguistic expressions, and used in the process of online meaning construction. Narrative

* Just a reminder of what is explained in the Reading Guide at the beginning of this thesis: the basics of the narrative-spaces framework are also discussed in Chapter 2 and 3, though with slightly different emphases. This is due to the fact that the chapters are written as separate papers, aiming at slightly different audiences.

spaces are each characterised by a particular set of features such as time, space, cultural norms, language spoken, or participants involved in either narration (narrators/focalisers), action (characters), or both (Dancygier, 2012: 35-37). These features are open to further elaboration by all kinds of local linguistic choices, such as usage of elements that coordinate viewpoints (for example complementation constructions or, as I argue, lexical units such as *allegedly* and *mistaken*), grammatical features such as tense or modality, usage of pronouns, typographical cues (such as quotation marks), etcetera.

Construction and elaboration of narrative spaces take mostly place through processes generally known as *framing* (Fillmore, 1985; for an overview see Cienki, 2007) and *blending* (Turner and Fauconnier, 1995; Coulson, 2001). The basic idea is that the background knowledge used by readers to interpret a text is structured in frames. Particular linguistic items used in the text activate a frame in its entirety, even if they relate to it only indirectly and in an unpredictable manner (cf. also the notion of "frame metonymy" as discussed in Fauconnier, 1997). Using this principle, writers of news articles can evoke rich meanings while providing only a few cues. For example, if a headline provides the information that a neighbour heard "non-stop shouting" coming from the Pistorius home during the hours before the shooting, nothing more needs to be said to evoke the frame of a *fight* between lovers, the sort of context in which one can imagine that things went *out of hand*, leading to a directed *attack* and hence a case of *murder* rather than a tragic accident (see also Section 4.3.6 below). In the current chapter I focus on a subcategory of this more general phenomenon, specific for viewpoint coordination and linked to a category of lexical items I refer to as viewpoint packages.

The general idea of blending is that two inputs with an established conceptual structure and content are integrated into an emergent *blended space* or *blend*. The blend has properties inherited from the inputs, as well as new structure and content of its own. Once a narrative space has been prompted and possibly further structured and enriched, it can as a whole be blended with another narrative space. The result of all the construction and blending processes is what Dancygier refers to as the *emerging story*: this is what a reader understands after having read and processed the text up until a particular point. The emerging story is thus a "moving end result".

Although Dancygier's framework is primarily designed to analyse narrative fiction, it fits our object of study in this chapter well. The ultimate question in the Pistorius case clearly is: was it an accident or murder? News reports provide a variety of micro-level cues that enable readers to put together a version (or rather, multiple competing versions) of the story covering the hours, days, or even weeks before the shooting took place. Clearly, the genesis of the story is different: instead of an author inventing it and "feeding" it gradually to the readers, the journalists themselves build their understanding of the case on various sources of facts and opinions. My focus here is on the linguistic choices these journalists make when subsequently feeding what they have understood to the readers of the newspapers. The narrative-spaces framework can be used to analyse how the journalists' linguistic choices at the micro-level of news reports relate to the formation of an increasingly rich and complex emerging story understood by the readers of the news.

One of the fictional examples Dancygier uses to introduce the concept of narrative spaces comes from Eggers' novel *A Heartbreaking Work of Staggering Genius* (2000). The main character Dave worries about having left his little brother Toph at home with a baby-sitter. While he is driving in his car, anxious thoughts cross his mind:

(2) I will come home and the door will be open, wide. The baby-sitter will be gone [...] Blood on the walls [...] a note to me [...] There will be a hearing, a trial, a show trial –

How did you come to meet this man, this baby-sitter?

We found a posting, on a bulletin board.

And how long did your interview of him take?

Ten, twenty minutes.

(based on Dancygier, 2012: 38, citing Eggers 2000: 126; italics in Eggers' original)

The main narrative space here is a period in Dave's life. Within this main narrative space, particular micro-level language phenomena prompt the construction and structuring of additional narrative spaces. Examples of such

phenomena are the choice of pronouns ("I", "you", "we"), choice of verb tense and modality ("will come", "will be", "did", "found"), and the use of italics ("*How did you...babysitter*"). The result is an emerging story that contains an increasingly rich and complex structure. Within the basic setting of Dave's life, the moment captured by the passage in (2) is a car ride to San Francisco. Within the setting of that car ride Dave imagines the scenario of coming home to where he left his little brother and finding blood on the walls, followed by an imagined trial in which he is being interrogated about the babysitter, realising (but this is not spelled out) that he did not find him through very reliable channels and did not take much effort to check on him either.

The emerging story at the end of (2) is indeed a blend of several narrative spaces. At the point of the interrogation, the story has parts of the structure of all these spaces: Dave is still in his car driving to San Francisco, and at the same time, through the layers of imagined events, there are his empty house with blood on the walls and the interrogation in court. However, readers will clearly not conclude that little brother Toph is "really" dead and Dave is "really" facing a trial. This is because they have sequentially read through the process in which the structure was built up one layer after the other, prompted by language phenomena that not only added content and structure, but also provided information about *how* this content and structure had to be integrated in the emerging story. The modal verbs at the beginning of (2) ("I *will* come home", "the door *will* be open", etc.) and other formal choices (such as the use of ellipses and enumerations, comma's, italics, etc.), along with the absence of any direct evidence of a crime (such as a phone call from the police that Toph was found dead), signal to the readers the status of the presented content: the blood, the note, the trial, etcetera, must be products of Dave's imagination. But what the readers *do* conclude from the fact that Dave is imagining doom scenarios, is something about Dave's overall mental condition: he probably has hysteric or paranoid tendencies, in particular when it comes to his little brother. However, overarching terms such as "hysteria" or "paranoia" are nowhere mentioned in the text. In other words: the constituents (imagined terrifying scenarios) are given, but the "whole" (hysteria) is omitted and left for the reader to construe. This is an instantiation of what I refer to as *compositional* complexity: the components that constitute the structure of the emerging story are all spelled

out in the text and each adds a single part of the total complexity, while the construction of an overall diagnosis of what is going on is left to the reader.

4.3.2 Compositional complexity: an example from the news

Similar to this, viewpoint layers are spelled out explicitly, hence complexity is construed compositionally, when a form of sentence embedding is used to coordinate perspectives. Consider again the sentence cited in the introduction, here repeated for convenience:

> (3) Athlete Oscar Pistorius **allegedly accidentally** shot dead his girlfriend at his house in Pretoria on Thursday morning, *Beeld.com* reported.
> (SAPA, 'Oscar Pistorius shoots girlfriend: report', 14 February 2013).

In the introduction I have stated provisionally that the first layer (*Beeld.com*) is related to the three others using "a form of indirect speech". However, in the literature there is no consensus on the question whether this should be considered a case of Indirect Speech (IS) or Free Indirect Speech (FIS). The expression in (3) as a whole is a specific syntactic pattern consisting of a reported clause followed by a reporting clause (underlined) on which the former is grammatically dependent. The clauses are not paratactically related, and the reporting clause is not grammatically complete without the reported clause (*"*Beeld.com* reported."). Consequently, the construction is one in which the reported situation is embedded in the point of view indicated in the reporting clause, and in this respect it is comparable to complementation constructions (although the different syntax also suggests semantic differences). Jeffries and McIntyre (2010: 89) would categorise it as IS, arguing that a sentence such as "The weather was nice, John said" only differs from "John said the weather was nice" in the order of the clauses, and the latter is clearly IS. Leech and Short (2007: 276) also mention "inversion", but at the same time have more of an eye for the differences when calling this pattern "Janus-like", "somewhere in between IS and FIS". Toolan (2006: 703) observes that the complementiser "that" cannot be inserted in preposed reported clauses and that they sometimes clearly require independent clause syntax ("Could he

accompany her home, he asked"; not *"He could accompany her home, he asked"). He therefore concludes that preposed reported clauses are more like prototypical FIS, and might even be considered FIS if it were not for the inquit-formula (for more discussion and examples see also Vandelanotte, 2009; 2012).

Given the presence of the inquit-formula (indicating a construal of the reported clause as in some way dependent), while at the same time agreeing with Toolan that this is clearly not a complementation construction, I propose to consider (3) as a relatively autonomous embedding construction, which I will refer to as *inquit-construction*. The important similarity between an inquit-construction and "classic" indirect speech using complementation, is that both coordinate two viewpoint layers that are spelled out in the text separately. Readers are cued to interpret one part of the sentence's content as being part of one layer (or narrative space), and another part as belonging to a different one. This is what makes it a form of compositional complexity: as in the Eggers-example, the constituting layers are spelled out in the text, whereas the overarching "whole" is not referred to explicitly. The grammar of English thus contains a family of constructions, including complementation and the inquit-construction, for cuing this kind of compositional complexity.

4.3.3 Holistic complexity: viewpoint packages

By contrast with compositional complexity, viewpoint packages introduce complexity holistically: the wholes are given in the text, while the underlying constituents remain for the reader to construe or unpack – if the context so requires. I argue that words such as *allegedly* and *accidentally* (bold in (3)) are holistic prompts to readers to imagine viewpoint layers that are not spelled out in the text. Vandelanotte discusses *alleged* as what he calls a "framing adjective", capable of "set[ting] something apart as belonging to a 'second-order' reality, viz. the reality of another's discourse" (2007: 360; referring also to McGregor, 1997: chapter 6). Other examples of framing adjectives are *so-called*, *pretended*, *purported*, or *supposed*. Vandelanotte suggests that "their shared reportative-evidential meaning can tentatively be glossed" as follows: "[framing adjective] X = stated by someone, but not the speaker, to be X" (2007: 368). Developing this line further, I suggest that Vandelanotte's "tentative gloss" reflects a *topology*, a

piece of conceptual structure evoked by the adjective. In the case of *allegedly* this conceptual structure looks in some respects similar to the one set up by a complementation construction or inquit-construction: one part of the related content must be attributed to the speaker and another to a third party. I will argue that such topologies implying coordination of viewpoints are not exclusive to framing adjectives; words that can function in this way can be found across several parts of speech. As said, the label that I propose for this type of words is *viewpoint packages*.

Looking at *accidentally* in this way, the following topology can be postulated: there is a particular outcome of an action, which differs from the one intended by an involved agent. In the actual discourse concerning the Pistorius case this topology is mapped onto situation-specific details, such as the athlete being in his house with his girlfriend at night and shooting her. When processing sentence (3) above, the reader will assimilate (through blending) the topology of *accidentally* with these details and take it that the athlete shot dead his girlfriend, but had not intended to do so.

In the case of *allegedly* it is given in the topology that a source different from the speaker asserts the information under the scope of this adverb. The details provided by the context in (3) only partly elaborate the topology and leave the identity of this other source open: it could be the police detectives or someone else who is in a position to make the claim that Pistorius (claims that he) shot dead his girlfriend accidentally. What content exactly is affected by *allegedly* depends on the interpretation of its scope: if only *accidentally* comes under the scope of *allegedly*, the result is "he shot her and it is said that this happened accidentally"; if the whole predicate of the reported clause comes under its scope, the result is "it is said that he shot her dead and that this happened accidentally".

What *allegedly* and *accidentally* have in common is that they entail an extra perspective layer from which the content they relate to is viewed, thereby modifying the way in which this content should be integrated into the emerging story. In the narrative spaces framework (as in general mental space theory), they would have the role of *space builders*. More precisely, they are linguistic cues that not just prompt any new space, but they prompt a *structured* space or *frame*, a space with a characteristic conceptual structure or topology. In Section

4.3.6 I will discuss this further, focussing on how readers assimilate the topology with local contextual details and integrate the result into the emerging story. Before that, two other issues will be looked at: potential context-specificity of *alleged(ly)* and ways in which compositional and holistic constructions of complexity are used in tandem to represent the entire thoughtscape underlying the Pistorius case in a press release published on the day after the shooting.

4.3.4 Alleged(ly): hedging or coordination of viewpoints?

It should be noted that *alleged(ly)* is commonly added in journalistic discourse related to juridical issues and criminal offence: news media can use it as a way of avoiding responsibility for anything that is still under consideration in court, where they would otherwise be liable to charges of slander or libel. This prompts the question whether *allegedly* is a conventionalised "hedge" associated with criminal and juridical reporting, rather than a more flexibly applicable way of coordinating viewpoints.

It seems possible that a viewpoint package becomes specialised for a particular context, thereby gradually losing the option of being "upacked" and worked out in terms of viewpoint layers, hence of coordinating viewpoints in a more flexible way. Vandelanotte's (2007) corpus research indeed suggests that the adjective *alleged* shows a degree of specialisation, reflected in frequent collocations with words referring to criminal offences and police investigations. In such contexts it normally "transfers" responsibility for a particular (phrasing of a) claim to an authority advancing the charge or dealing with the investigations in a legal case. The example given in Vandelanotte's discussion (2007: 363) illustrates this particular use:

> (4) In New Jersey today, a jury will hear closing arguments in the trial of four young men accused of raping a mentally retarded woman. The **alleged** rape took place in the suburban town of Glen Ridge, New Jersey, four years ago this week.

Assessment of 100 randomly chosen instances from the British National Corpus (BNC) showed that the adverb *allegedly* is also commonly used in this way, but

by no means exclusively.[60] Ample instances (in the current sample at least 20) can be found where the external source of viewpoint referred to by *allegedly* is not some sort of legal authority, but, for example, a particular tradition or history (e.g. "Simon Peter [...] on whom Jesus allegedly founds his church"; BNC, EDY), a source of unconfirmed or even questionable authority (e.g. "one tries to sell [...] smear in a bottle allegedly from the great Madonna herself"; BNC, CBC), or an entity specified in the direct context (e.g. the protesters in "the protesters' vociferously expressed and allegedly 'sincere' ideals"; BNC, HTP); it is unclear from my sample whether the use of quotation marks is indicative of this type of use).

Although an important effect of the insertion of *allegedly* in the press release cited in (3) is the relegation of responsibility for the content under its scope to an external party, I suggest to see this viewpoint package here not primarily as a conventionalised hedge bound to a particular context. Rather, I argue that it is a lexical item cuing the coordination of two viewpoint layers in a way not very different from the working of complementation and inquit-constructions. There are differences in what is in general left implicit or realised "on stage". I expect that this reflects a more general pattern: in the case of complementation or inquit-constructions the party responsible for the external viewpoint appears to be given by default (although impersonal constructions are possible: "It is claimed that..."); in the case of viewpoint packages this party can more easily be left implicit or "off stage" (although it can be elaborated in the context: see the BNC examples cited above). Future (corpus) research will have to shed more light on such differences in emphasis, focussing in particular on how responsibility is distributed over the speaker and the external source of viewpoint, and the degree to which both are on or off stage (see also Langacker, 1990; Wierzbicka, 2006; Vandelanotte, 2009, 2012; Dancygier, 2012a).

[60] Using SketchEngine (http://www.sketchengine.co.uk; see Killgarif et al., 2014) a random sample of 100 instances was drawn from the 1039 instances of *allegedly* in the British National Corpus (BNC; 112,181,015 tokens in total). The BNC is distributed by Oxford University Computing Services on behalf of the BNC Consortium. All rights in the texts cited are reserved. For information, licensing conditions, and use of the text identifiers (the three-letter codes EDY, CBC, and HTP in my citations) see http://www.natcorp.ox.ac.uk.

4.3.5 Representing the thoughtscape

When a larger excerpt of the press release starting with sentence (3) is considered, it becomes clear that multiple means of viewpoint coordination are being combined and mutually embedded to allow readers to form an understanding of the full complexity. Consider the second sentence of the press release:

> (5) He had **mistaken** her for a robber, the *Afrikaans daily* reported on its website.
> (SAPA, 'Oscar Pistorius shoots girlfriend: report', 14 February 2013)

In this sentence, *mistaken* functions as a viewpoint package picking up on the earlier *accidentally*: it entails an extra viewpoint layer by implying that at a certain point the athlete *believed* that something was the case, whereas in actuality it was not. Another inquit-construction can be found: "the *Afrikaans daily* reported [...]" has scope over the clause containing *mistaken*. The past perfect "had mistaken" (as opposed to simple past "mistook") is also involved in viewpoint coordination and has an effect similar to that of *allegedly* in (3): the responsibility for the reported mindstate of the athlete is not (fully) attributed to the *Afrikaans daily* but relegated to a third party. So, example (5) contains a combination of three different means of viewpoint coordination: sentence embedding (here an inquit-construction), usage of the tense system (here past perfect), and usage of a viewpoint package (*mistaken*). However, as announced at the beginning of this chapter, here I focus on a comparison of viewpoint coordination effected by viewpoint packages (holistic) on the one hand and grammatical patterns such as complementation or inquit-constructions

(compositional) on the other, and refrain from detailed analysis of the viewpoint effects of tense, modality, passive voice, or negation.[61]

The next sentence of the press release exhibits not only a combination of viewpoint packaging and embedding, but a form that may be called a hybrid of the two:

> (6) Police spokeswoman Captain Sarah Mcira **confirmed** she [Reeva Steenkamp, MvD] was shot in the arm and head.
> (SAPA, 'Oscar Pistorius shoots girlfriend: report', 14 February 2013)

As a whole, (6) has a form not very different from (3) and (5): it features a reported clause, this time in a complementation construction, attributed to the viewpoint of a specified source. However, a different verb is used: *to confirm* rather than *to report*. Whereas *to report* does not automatically imply information about perspectives other than that of its subject, *to confirm* entails the assumption that the content of the reported clause it introduces was already claimed to be true by someone else; or possibly: that this content is common knowledge. More precisely: for a speaker A to *confirm* proposition p for addressee B, means that A states p and implies that p has been claimed before, and that it is accessible to B that p has been claimed before. In terms of Clark's (1996) view that will be detailed in Chapter 5: p must be part of the *common ground* of A and B. Viewed this way, *to confirm* can be said to be a lexical unit with a topology containing an extra viewpoint layer from which the related content is viewed, and therefore qualifies as a viewpoint package. As a result, the complex of viewpoints involved in (6) can be depicted as follows:

[61] Throughout various traditions of linguistic research, attention has been drawn to the way in which a range of linguistic phenomena are involved in managing viewpoint. For a volume discussing a wide variety of linguistic means for viewpoint coordination, see Dancygier and Sweetser (2012). More examples, among many others, are the description of the disjunctive function of particular adverbs (*allegedly* would come under this category) in Functional Grammar (see Pinkster, 1990: 4, 32ff; Greenbaum, 1969). Wierzbicka (2006, esp. chapter 7 and 8) usefully analyses semantic distinctions between a number of reporting verbs and "epistemic adverbs" (also including *allegedly*), in the context of the supposed English rationalist cultural norm of epistemic caution rather than that of the construction of complex thoughtscapes. For an account of perspectives and the aspectual adverbs *still* and *yet*, see Ter Meulen, 2003.

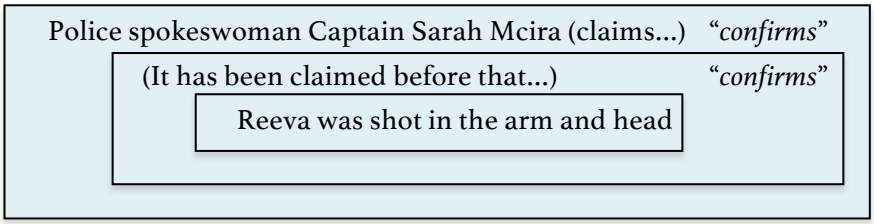


Figure 2

To round off this part, I will take the discussion of how complexity is constructed in the news reports back to Dancygier's narrative spaces framework.

4.3.6. Viewpoint packages, the emerging story, and (de)compression

In the Eggers example cited in (2) there were various linguistic cues signalling to the reader that the empty house, the blood on the walls, the trial, etcetera, were not as such to be integrated into the emerging story. All these crime-related details were embedded in viewpoint layers of Dave's imagination. Instead of concluding that Dave's brother is in severe danger, or perhaps even already dead, readers integrate into the emerging story a more general conclusion: that Dave has hysterical tendencies. This hysteria could be considered a holistic (or *compressed*; see Dancygier, 2012: 100-102) version of the viewpoint layers spelled out in the text, which is not mentioned as such but left implicit for the reader to construe. As we have seen, in some parts of the Pistorius reports we find the precise opposite: here, several of the constituent viewpoint layers are not spelled out in the texts, whereas their holistic counterparts are mentioned: the viewpoint packages *allegedly*, *accidentally*, and *mistaken*. This gives rise to the question whether readers at any point "unpack" (or *decompress*) the packaged structure in their minds; or in other words: whether they in some way construe the separate viewpoint layers contained in the packages.

Part of this question can be answered by looking at the extent to which news items following up on the first headlines build on the packaged viewpoint

layers. Consider the following two excerpts from a news article that was published one week after the shooting:

> (7) Mr. Pistorius told the court [...] he heard a strange noise coming from inside his bathroom, climbed out of bed, grabbed his 9-millimeter pistol, hobbled on his stumps to the door and fired four shots.
> (International Herald Tribune, 'Testosterone reported at home of Pistorius', 21 February 2013)

> (8) Prosecutors said [...] that Mr. Pistorius was calm and had the presence of mind to strap on his prosthetic legs, walk to the bathroom door and open fire as Ms. Steenkamp cowered behind it.
> (idem)

These two possible versions of what could have happened before Pistorius fired his gun present details that are intended to take up either on the scenario of an *accident*, in which the athlete *thinks* he is shooting at a burglar ("strange noise", "hobbled on his stumps"), or on that of *murder*, in which he *knows* he is shooting at his girlfriend ("presence of mind to strap on his prosthetic legs", "walk", Steenkamp "cowered" behind the door). Unless readers have some representation in their minds of these two competing scenarios and the mindstates appropriate to each, it hardly makes sense for newspapers to provide such details without elaborating on their relevance. It may be noted that there is indeed no need for newspapers to do this: details as cited in (8) and (9) can be (and are indeed widely) published without explicit attempts to place them in the story as a whole.[62]

More examples can be found; for instance: by presenting evidence that an argument took place just before the shooting, the prosecutors imply that the

[62] Note that the reports are highly redundant across various newspapers and other media. A week after the shooting, several pieces of information had been added to the "canon" represented in virtually all reports, including witnesses who claimed to have seen light or heard voices in the Pistorius home, small fragments of Pistorius' account of the events to the court, and small fragments of what the police detective leading the investigations has reported (e.g. that he saw a substance which could have been testosterone in the athlete's bedroom). They are often followed by a brief summary of news coverage of the case so far, but no explicit explanations of how they fit together are provided.

athlete *could have known* that Reeva was inside the bathroom when firing the gun:

(9) A witness heard "non-stop shouting" in the home of South African athletics star Oscar Pistorius shortly before his girlfriend was shot dead, the detective leading the murder investigation said on Wednesday.
(Reuters News, 'Witness heard 'non-stop shouting' before Pistorius shooting', 20 February 2013)

(10) While Mr. Pistorius had said the house was dark, the prosecution cited a witness as saying a light had been switched on when the first of four shots was fired. The witness heard a gunshot, then the sound of a woman screaming, then more shots.
(International Herald Tribune, 'Testosterone reported at home of Pistorius', 21 February 2013)

These details are then again challenged by the athlete and his lawyer: the untrustworthiness of these testimonies makes it impossible for the prosecutors *to know* whether or not Pistorius *could have known* that Reeva was in the bathroom:

(11) But the defense disputed that testimony, saying the neighbor who claimed to have overheard an argument in Mr. Pistorius' home in fact lived 600 yards, or about 550 meters, away.
(idem)

In order to be able to integrate such details as presented in (7)-(11) into the emerging story in a meaningful way, readers must already have a particularly structured representation of the thoughtscape in mind. At the very minimum, this representation must comprise the distinction between the two competing scenarios (accident or murder) and several of the viewpoint layers mediating

between the reader and the "actual past event" of the shooting.[63] A schematic depiction can be drawn as follows:

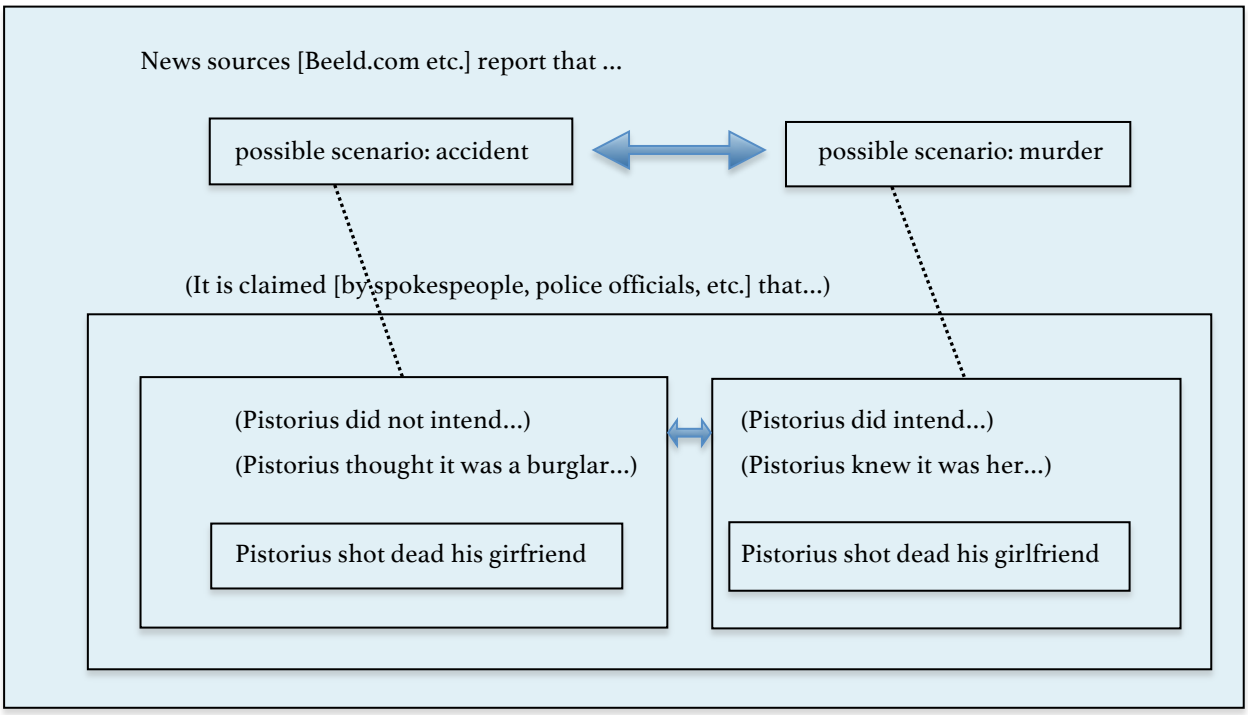


Figure 3

Figure 3 illustrates once more that the claim that Pistorius killed his girlfriend is embedded in an array of viewpoint layers. As stated above, these layers can partly be understood as the result of newspapers relegating responsibility to external sources, such as authorities and other media (see Section 4.3.3 and 4.3.4). However, another reason for their presence in this particular case is that the *actual past event* cannot be retrieved: no one, except Pistorius, knows what "really" happened. As a consequence, reports appearing on the first day after the incident are confined to presenting *possible views* on what has led to the known outcome: that Reeva Steenkamp was shot dead. This is precisely what *accidentally* and *mistaken* do in (3) and (5): they prompt the viewpoint layers of

[63] Cf. Dancygier's (2012) analysis of the opening passage of Margaret Atwood's *The Blind Assassin* and her Figure 4.1 in particular (89-91). There is an important parallel (besides several differences) between the Pistorius case and Dancygier's example: both start out with two alternative scenarios of "what happened", expressed in various viewpoint layers which provide the basis for the elaborations that follow.

Pistorius *thinking* it was a burglar and hence *not intending* to shoot Steenkamp, and arguably at the same time their alternatives of him *knowing* she was in the bathroom and hence *intending* to shoot her—after all, it is widely accepted that a negated proposition also entails its positive counterpart (see Fauconnier, 1997; Sweetser, 2006; Dancygier, 2012a).[64] In this way, the reports and press releases appearing on the first day after the shooting provide what Dancygier calls a "proleptic 'summary'" (2012: 90): they prompt a particular configuration of narrative spaces that persists as a fundament as further elaborations follow.

Regarding the linguistic cues involved in building up this configuration it can be observed in (3), (5), and (6) that sentence embedding is used to coordinate the viewpoints of various sources reporting on the incident. These are the "outer" layers of the schema in Figure 3: the viewpoints of news media and the authorities and their spokespeople. As a result of this choice of linguistic form of viewpoint coordination, the constituting viewpoint layers are presented explicitly in the text (complexity is constructed *compositionally*). By contrast, "inner" layers of the schema, i.e. the representation of the athlete's perspective at the moment of pulling the trigger, completely relies on viewpoint packages until more direct information becomes available several days after the shooting. This pattern is consistent throughout the twenty articles and press releases of the first day after the shooting I have looked at.[65] In general it is thus only on the basis of words such as *accidentally* and *mistaken* that readers can build an understanding of the athlete's possible mindstates when he fired his gun. Given that newspapers publish details following up on precisely these mindstates without further introduction, we may conclude that readers must have some form of access to the packaged (or compressed) viewpoint layers when the context so requires.

[64] This picture is consistent throughout twenty articles and press releases of the first day after the shooting I have looked at: fifteen refer to the shooting with *mistaken*, *mistook*, *mistaking*, *a mistake*, *accidental*, or *accidentally*. Three (of those fifteen) also speculate on the possibility of murder explicitly; the remaining five mention neither of the two alternative scenarios in explicit terms, but arguably imply both. The reports published on 14 February 2013 were retrieved on 21 February 2013 using Lexis Nexis (http://academic.lexisnexis.nl).

[65] The only exception being *Asian News International*: "Pistorius [...] has allegedly shot his girlfriend to death in the early hours of Thursday morning after thinking she was an intruder, reports claim".

Put differently: it seems that readers do not unpack (or decompress) packages by default—a package may usually be closed and taken on board holistically—but when necessary, the implied viewpoint layers can be accessed. This leads to a final core aspect of viewpoint packages. Consider the word *mistaken*: thanks to its holistic nature it does not automatically necessitate reflecting on a whole series of assumptions about "who knew what at which moment in time". Still, it is easy to prompt assumptions about the distribution of such information over minds by providing specific contextual details along with it. This has clear communicative, and conceivably also cognitive, advantages: constructing the viewpoint complexity underlying a situation compositionally every time it needs referencing (e.g. by using sentence embedding) is possible, but often unnecessarily ponderous. Therefore, in contexts where space and time are limited (as is the case in news reports and headlines), holistic introduction of complexity using viewpoint packages appears to be a preferable option. The difference between the actual newspaper quotes in (3), (5), and (6) and the unpacked thoughtscapes as explicated in the introduction of this chapter ("news media *suggest* that various sources *report* that the athlete *claimed*…etc.") and depicted in Figure 1-3 testifies to that: whereas the first will be perceived as natural and everyday formulations, the latter present the same information in the form of a layered structure that cannot be absorbed at a glance.

4.4 Conclusion and discussion

4.4.1 Future directions: processing and acquisition

In the past two decades, a fair amount of attention in cognitive and evolutionary psychology has been devoted to the cognitive challenges humans face in their social environments. In order to function well socially it is of great importance to be able to reason about mindstates of others, that is: assess what they think, believe, intend, desire, and so on. As a result, primate social life is

highly demanding of cognitive resources.[66] Dunbar (e.g. 2003) suggests that humans typically form social networks of around 150 individuals. In his view, to maintain such a network, humans must be able to process what has been termed *multiple-order intentionality* up to five or six orders. Examples of such assumed processing tasks involving four orders (12) and five orders (13) are:

(12) Jenny *hoped* the greengrocer *believed* the chemist *had wanted* to give Emma a job. (Stiller & Dunbar, 2007)

(13) (a. Connie *knew* that John *suspected* that Pete *thought* that Sheila *hoped* that John would ask her out. (idem)

Note that these statements show similarities to the phrasing of the thoughtscape underlying the Pistorius case in Section 4.1 of this chapter and to the layered structure depicted in Figure 1-3. Various studies have used such statements in multiple-choice questions to test a participant's abilities to process multiple-order intentionality (see Chapter 6). Where participants generally make few mistakes in questions covering up to fourth-order intentionality, error rates increase quite drastically in questions that involve fifth- or sixth-order. This has led to the suggestion that humans face a natural limit at this point (Kinderman, Dunbar, and Bentall, 1998; Launay et al., 2015).

However, in Chapter 2 I have argued that the way in which situations involving multiple-order intentionality (which I call thoughtscapes) are represented in discourse, greatly influences the actual performance by humans in processing such tasks. Consider Shakespearean drama: by the end of act II, the audience of *Othello* has to *understand* that Iago *intends* that Cassio *believes* that Desdemona *intends* that Othello *believes* that Cassio *did not intend* to disturb the peace. When represented like this, using an embedded sentence structure, this is highly opaque, whereas it is beyond doubt that the play has been

[66] This is true throughout the entire primate world, where social groups are structured around dyads: every individual has a personal relationship to some or all of the other individuals in the group. Having a personal relationship with someone, involves keeping track of a lot of knowledge about this individual, including what this individual knows about other group members and oneself (e.g. David-Barrett & Dunbar, 2013). This idea is central in work on the *social brain hypothesis* (e.g. Dunbar, 2003; Byrne & Whiten, 1988). See Chapter 1, Section 1.3 for more details.

understood and appreciated by innumerable different audiences for several centuries. I have suggested that in *Othello* and comparable stories, certain "expository strategies" characteristic to narrative facilitate the effective representation of complex networks of mindstates, thereby alleviating the reader's or spectator's burden of processing such complexity cognitively. In Chapter 3 I have come to a similar conclusion when analysing excerpts from several novels.

In this chapter I have suggested that viewpoint packages serve to communicate complex thoughtscapes efficiently and naturally. They allow readers to take on board certain parts of a layered structure holistically, while the underlying complexity can be unpacked (or decompressed) if the context so requires, but need not be otherwise. Extending this point to the domain of cognitive processing, I suggest that, in psychological terms, viewpoint packages have *Gestalt*-like properties: they can be used as holistic items, while their constituent components remain accessible (Lakoff, 1977; see also Gigerenzer, Hertwig, & Pachur, 2011; Humphrey, 1924). As such, viewpoint packages could provide crucial scaffolding for the cognitive handling of complex thoughtscapes: given constraints of working-memory and processing, working with packages allows humans to engage in more complex patterns of reasoning than working with substantively identical, but non-packaged primitives (Beekhuizen and Van Duijn, 2013). This is an alley, I propose, that should be further explored in the future. It would fit with the broader idea that human cognition relies crucially on "tools" obtained through social learning and potentially accumulated through cultural evolution (see Tomasello, 1999; Dennett, 2000). Viewpoint packages can be considered such tools: when we learn to use them in language, their representational power opens new worlds of possibilities for our mental and communicative activities, not replacing but complementing existing ones.

Finally, it is worth adding that there is another interesting connection between viewpoint packages and complementation constructions within the realm of language acquisition. Diessel & Tomasello (2001) have demonstrated that the earliest items that *look like* instances of complement taking predicates (thus of complex syntax) in children's utterances actually have the status of formulaic items that mark subjectivity or illocutionary force (*I think …*, *I wish …*,

Look …). Thus, they may at that stage be regarded as a kind of viewpoint packages in the sense of this paper. In a developmental perspective, they differ from the other viewpoint packages discussed here in that they are relatively simple (but this is a matter of degree), and that they form the starting point of the growth, in due course of development, of a network of complementation constructions that also includes productive abstract patterns (cf. Verhagen, 2005: 110). The latter is not the case for the English adverb *allegedly*, but the process of the acquisition of such items and the social cognitive capacities underlying it (both in initial and in later acquisition) may be assumed to be similar, if not the same. I leave it to future research to elaborate and test this intriguing possibility.

4.4.2 The refined definition of viewpoint packages

Throughout this chapter I have looked at newspapers reporting on the Pistorius case, finding that already on the first day after the shooting a complex thoughtscape was covered by headlines and news articles. This complexity (depicted schematically in Figure 3) was partly cued compositionally, with the constituting layers in the text, and partly holistically, using words such as *alleged(ly)*, *accident(ally)*, and *mistake(n)*. I have proposed to label such words "viewpoint packages" and provided a preliminary definition, which can now be refined: a viewpoint package is a single lexical item that entails at least one implicit viewpoint layer. This (or these) layer(s) allow for the attribution of (parts of) the content to the viewpoint of particular discourse participants that can remain unspecified (as was the case with *allegedly*) or identified in the context (for example "the athlete" with *accidentally*). In this way, a viewpoint package can be used to regulate the (perception of the) amount of responsibility that is taken by the speaker and other involved parties exhibiting intentionality.

In addition, a viewpoint package lets language users take on board complexity holistically, while underlying layers remain accessible: it can be decompressed if the context so requires. This makes it an efficient tool for communicating complex thoughtscapes, and possibly also for processing them cognitively.

Chapter 5

Language and joint intentionality: reflecting on orders of intentionality is the exception, not the default when communicating[*]

5.1 Introduction

In section 1.2.2 of Chapter 1 I have distinguished the three main roles of language in relation to mindreading, as used in this thesis. Language can *represent* mindstates and the relationships between them (first role), either formally, in propositions of the form "A thinks that B intends that C...etc.", or naturally, using a mix of different linguistic elements capable of viewpoint coordination. Various types of these linguistic elements were discussed throughout the previous chapters. Going from the smallest to the largest level of analysis, these were: lexical items such as the viewpoint packages "allegedly", "accidentally", and "mistaken", grammatical constructions such as complementation and the inquit-construction, the patterns of focalisation and reported speech and thought (STR) that coordinated the different perspectives presented in Woolf's *Mrs Dalloway*, and the expository strategies of narrative that aided the audience in understanding the complex thoughtscape underlying Shakespeare's *Othello*. What all of these linguistic elements had in common was that they provided conventionalised ways in which speakers of English could represent mindstates and the relationships between those in discourse—from a single belief held by one person up to an entire thoughscape.

[*] Versions of this chapter, especially Sections 5.3 and 5.4, were presented at the 47th Annual Meeting of the Societas Linguistica Europaea (SLE) in Poznan, Poland, 11-14 September 2014, and at the Perspective Project Kick-off Meeting in Nijmegen, The Netherlands, 17 November 2014. See the Reading Guide for more details.

In this way, it was argued that natural languages support not only efficient communication of mindstates and thoughtscapes, but seem also capable of providing *support for cognition* (second role). When looking at one particular usage event, language can provide a form of short-term "online" support: I have argued that the way in which mindstates and their mutual relationships are represented linguistically can execute strong influence on the ease or difficulty with which they could be processed (think of the expository strategies making a complex thoughtscape manageable and viewpoint packages conveying multiple intentional relationships at once in a holistic way). When looking at the longer term, language users somehow internalise ways in which language makes mindstates and their relationships insightful, which can account for what some researchers have referred to as "implicit support" for mindreading: I have discussed research suggesting that various aspects of language usage, once mastered, work as scaffolding, conceptual underpinning, or training for our intentional reasoning skills. For example, experimental evidence suggests that children aged 3-4 who were for a while exposed more intensively to embedding constructions and perspective-shifting discourse, pass false-belief tests earlier, presumably because their general "thinking repertoire" got enhanced when they learned to master particular grammatical patterns (Lohmann and Tomasello, 2003; Milligan et al., 2007). Also it has been suggested that stories in all their different appearances, ranging from the day's latest gossip or a myth told around the campfire to an award-winning novel, help both children and adults to develop and sharpen up their mindreading skills over time (see Chapter 1, Section 1.2.2; Chapter).

Language was also argued to be itself heavily *dependent on* our mindreading abilities (third role). In the current chapter, this dependency will be investigated in more detail. According to researchers such as Sperber (1994; 2000) and Scott-Phillips (2015), it takes the capacity to reason at four or five levels of intentionality to exchange even very basic utterances.[67] This position is intuitively controversial: if language is naturally capable of representing

[67] I agree with both Scott-Phillips and Sperber on many points regarding language and meaning. However, there also is an important issue on which I disagree: the way they construe the relationship between linguistic interaction (or, more generally, "pragmatic competence") and multiple-order intentionality, which they refer to as "recursive mindreading" or "recursive metarepresentation". This issue will be central in this chapter.

complex thoughtscapes, does that mean that an addressee of a short story involving, say, four mutually linked perspectives essentially has to deal with a total of eight or nine intentional states, four or five from the communicative situation plus four from the story? And if so, how can this be unified with evidence that dealing with multiple intentional states is cognitively demanding, or, for that matter, with the claim that humans can deal with *at most five levels* of intentionality reliably (see Chapter 1, Section 1.2.1)? What is the role of linguistic and narrative elements capable of viewpoint coordination, such as the viewpoint packages discussed in previous chapters? And how does all of this affect our evolutionary story?

These questions lead up to the objectives of this chapter, which has two main parts. First, as announced in Chapter 1, I will contest the claim made by Sperber and Scott-Phillips. Using the concepts of *common ground* (Clark, 1996) and *joint cognition* (Hutchins, 2006; Verhagen, 2015) I will argue that only in exceptional cases do we need to bother about any layers of intentionality. Regarding some aspects I will be relatively brief in my analysis, and refer to existing work or point out opportunities for future research. Other aspects, however, will turn out to be closely tied to points made in the previous chapters, and be elaborated in full detail. An important role will be played by the *ratchet effect*: linguistic items "store" communicative experience of generations of language users on which every new generation can build.[68] This, then, leads to the second part of this chapter: providing an *integrated conceptual model* for analysing the particular class of linguistic elements central in this thesis so far, namely: elements capable of viewpoint coordination in discourse. After the model has been introduced and explained, I will briefly explore some of its consequences for our evolutionary story.

[68] The term "ratchet" is taken from Tomasello (1999). My usage of it here is compatible with his, however, I apply the idea more specifically to linguistic items whereas Tomasello speaks about cultural conventions more broadly.

5.2 Association, ostension, and shared intentionality

Throughout the literature, it is quite generally recognised that human communicative interaction "as we know it" requires some form of mindreading on behalf of both interlocutors, irrespective of whether we use language, gestures, facial expressions, or any other means to get our messages across (see e.g. Verhagen, 2005; Levinson, 2006; Tomasello, 2008). The usual argument is that signallers have to design their communicative behaviour such that their particular audience will be able to infer from it what they mean, and addressees have to reckon why a signaller picked out this particular behaviour—both these tasks entail a degree of understanding of the other's mindstate. However, on top of this, some researchers have made a case for why human communication cannot succeed just by virtue of basic mindreading competencies. Indeed, Scott-Phillips makes this point explicitly: what most linguists and philosophers of language have failed to appreciate, according to him, is that sophisticated intentional reasoning skills including "recursive mindreading" are a prerequisite not only for the successful execution of communication-as-we-know-it, but also for such communication to evolve and develop at all (2015: 68-69). The argument thus has two components: it deals with the question "Which mindreading skills enable interlocutors to take part in communicative interaction as we know it today?" ("synchronic"), and with the question "Which mindreading skills were necessary for the emergence of such a form of communicative interaction in the first place?" ("diachronic"). It should be noted that these two components are not always brought forward and supported separately by Scott-Phillips, but in this chapter I find it useful at several points to keep the synchronic and the diachronic stories apart.[69]

[69] Note that this is a different divide than the "classic" one between ontogeny and phylogeny (Tinbergen, 1963). Here I mean not "development over a lifetime" versus "development over evolutionary time", but "the working of communication as it is now" versus "the emergence of such communication over time". In fact, my notion of "synchronic" is closest to Tinbergen's question of the proximate mechanism, whereas what I call "diachronic" covers both his developmental and evolutionary questions.

5.2.1 Scott-Phillips' two models of communication

The starting point of the idea advocated by Scott-Phillips goes back for a large part to Sperber (1994; 2000) and Sperber and Wilson (1995; 2002), and ultimately has its roots in what could be called the "pragmatic turn" in linguistics and philosophy of language that began with the second half of the twentieth century.[70] It sets human communicative interaction, whether or not involving language, apart against other communication found in nature by arguing that it is, at its core, not a system of "coding-and-decoding" information. A coding-and-decoding system can be found in (to follow Scott-Phillips' example) grasshoppers producing six different signals associated with six different states of the grasshopper world: "I would like to make love", "You are trespassing my territory", "How nice to have made love!", and so on (2015: 5, citing Moles, 1963: 125-126). Various forms of code-system communication can be found throughout the primate world, including in humans, ranging from olfactory cues (smell) guiding behaviour of newborns, to spontaneous emotional vocalisations working as alarm calls, and, potentially, (Duchenne) laughter signalling social solidarity (see Scott-Phillips, 2015: 5-6; Burling, 2005). Even though such code systems need not be fully deterministic or, for that matter, trivial, the primary mechanism linking signals to messages is *association*: every signal type stands for one particular meaning type (or, if a code is probabilistic instead of deterministic: a class of meaning types). However, association falls short of explaining human communicative interaction, given that there are many ways in which we can express a particular meaning, while at the same time all of our expressions can have multiple different meanings. The example given in Section 1.2.2 of Chapter 1 was that of someone saying to a friend "hey, there is Ann", which could mean "all right, we can go inside", "let's go somewhere else", "what a coincidence", and so on, depending on the circumstances. The same goes for non-linguistic communication: if we raise a full glass of beer in the air while seeking eye contact with someone who also holds a full glass, this probably means "cheers!", while it can also mean

[70] Sperber has developed his insights on the basis of Grice (1957), although Grice was not interested in evolution. For overviews covering also the important contributions made by Austin (1962), Searle (e.g. 1969), and Wittgenstein (e.g. 2006 [1953]) see Hacker (1986: esp. chapter 6-11) and Keller (1995).

"thanks!" if the other person is the one who just paid for the round, or "do you want one as well?" if the other is holding an empty glass.

Neither the utterance "hey, there is Ann" nor the behaviour of raising a glass and seeking eye contact stand for all of these meanings in the sense of the code model: there are no one-to-one associations. There is a different system at work that forms the basis for the production of meanings, which Scott-Phillips describes as the "ostensive-inferential" model of communication (2015: 7-13; see also Sperber and Wilson, 1995). According to this model, signallers have the intention to alter an addressee's mindstate or behaviour in some way. They provide particularly designed evidence for this, thereby enabling the addressee to draw the right inferences. This evidence can take the form of a string of words, but could, depending on context and desired effects, just as well be a set of gestures, facial expressions, or any other behaviour, as long as it is in some way *ostensive*: it has to be possible for the addressee to infer not only *what* the signaller wants her to understand (referred to as the "informative intention"), but also *that* the signaller is trying to communicate this in the first place (called the "communicative intention"). As an illustration, consider the example he borrows from Sperber (2000):

> Mary is eating berries. She wants Peter to know that she thinks that the berries are very tasty, so she eats them in a somewhat exaggerated, stylized way, and pats her tummy as she does so. This reveals two things to Peter: (i) that Mary thinks the berries are tasty (this is the content of her informative intention); and (ii) that Mary wants to communicate this fact to Peter (this is the content of her communicative intention). If Mary simply ate the berries enthusiastically, but did not do so in a stylized or exaggerated way, Peter would still be able to infer that they are tasty, but not because Mary had expressed either an informative or a communicative intention. There would be no communication in that case.
> (Scott-Phillips, 2015: 9)

In other words, given that there is no fixed set of signals associated with particular messages in this case, Scott-Phillips (along with Sperber) suggests that each signal must in principle first be negotiated *qua signal*—a process that

is explained by the ostensive-inferential model of communication, but not by the code model (see also Stolk, 2014, for discussion and an experimental approach). Ostension and inference are thus the basis of human communication, according to Scott-Phillips. On top of this, he argues, there is also a code at work: the conventions of a language provide global links of linguistic forms to certain meanings. In this way, ostension and inference make human communication possible in the first place, and the linguistic code makes it even more powerful (2015: 15-17).

The two different models of communication require quite different skills on a cognitive level.[71] In principle, the code model only requires a "glossary" listing all signals and associated meanings (which can be as simple as with the grasshoppers, but also more complex). This can be a genetically inherited glossary, but the capability to develop such a glossary through associative learning can also do the job. The ostensive-inferential model, by contrast, requires a great deal of flexible reasoning abilities, including mindreading. In order to design the right evidence for their intended meaning, signallers need not only take into account the context (where and when the communication takes place, who is present, etc.), but also what their addressees (already) know and believe about the topic and context. Addressees, in turn, must factor in what they think the signaller believed about them, the topic, and the context when designing the signal, in order to make the right inferences. Both interlocutors must thus be able to reason about contextual factors, including the other's intentional states, for ostensive-inferential communication to be possible.

I support the distinction between the code model and the ostensive-inferential model and agree with the analysis that the requirements on the cognitive level are the ability to form associations in the case of the code model, whereas flexible reasoning abilities including mindreading are needed in the case of ostensive-inferential communication. Yet this is where the debate begins: I disagree with the amount and complexity of the mindreading Scott-

[71] My aim in this section is clearly not to provide *exhaustive* lists of what is required for communication on a cognitive level. Rather, I will highlight important differences between the kind of cognitive structure needed for the code and inferential models to work (see also Scott-Phillips, 2015: 64), and in 5.2.3 I will do the same for my alternative communication model.

Phillips and Sperber consider to be necessary. In what follows I will argue that they misconstrue the complexity needed *in theory*. On top of that, I will argue that *in practice* we hardly need any mindreading at all for successful communicative interaction, by discussing various mechanisms that save interlocutors from cognitively taxing mindreading efforts.

5.2.2 Cognitive requirements of ostensive-inferential communication

As said at the beginning of the previous section, Scott-Phillips explicitly makes the point that many others across the literature agree that *some* mindreading is needed for human communication, but that its exact role and complexity are rarely spelled out. In order to get a grasp on this, he sets up the following argument, using a series of different scenarios taken from Sperber (2000):[72]

> Scenario one. Mary is picking and eating berries. She does this because the berries are edible.
>
> *Scenario two.* Mary is picking and eating berries. Peter is watching her, and hence forms a belief about the edibility of the berries. Here, *Peter believes$_1$ that the berries are edible* (because otherwise Mary would not be eating them). Note that Mary may or may not know that Peter is watching. Whether she does or not, it makes no difference to her intentions or behaviour.
>
> *Scenario three*. Mary is picking and eating berries. Peter is watching her. Mary knows Peter is watching her, and she wants him to believe that the berries are edible. So: *Mary intends$_1$ that Peter believes$_2$ that the berries are edible*. Here, note that Mary's behaviour is identical to her behaviour in scenarios one and two. All that has changed is that in scenario two Mary informed Peter about the edibility of the berries only incidentally [...] whereas here she does so intentionally – and she can satisfy this intention (that Peter believes hat the berries are edible) simply by picking and eating berries. She need not and does not do anything more than this. Mary's intention is an informative intention.
>
> *Scenario four*. Mary is picking and eating berries. Peter is

[72] Note that Scott-Phillips uses numbers in subscript to indicate orders of intentionality: "Mary intends$_1$ that Peter believes$_2$ that...".

> watching her. Mary knows Peter is watching her, and she wants him to believe that the berries are edible. Furthermore, Peter knows that Mary knows that he is watching her and, for whatever reason, he has reason to believe that she would like him to believe that the berries are edible. Correspon-dingly, *he believes$_1$ that she intends$_2$ that he believes$_3$ that the berries are edible*. Mary, however, does not know that Peter believes this. After all, she has not yet made her intention manifest to Peter. Indeed, Mary's physically observable behaviour is the same as it is in scenarios one, two, and three. As yet, she has not picked berries in a way that signals to Peter that her behaviour is intended to be informative. She has not yet signalled signalhood. All that is different between this and scenario three is that here Peter believes, correctly, that Mary has an informative intention.
>
> *Scenario five*. Mary is picking and eating berries. Peter is watching her. Mary knows Peter is watching her, and she wants him to believe that the berries are edible. Furthermore, Peter knows that Mary knows that he is watching her, *and* Mary knows that Peter knows this. As such, when she eats the berries, *she intends$_1$ that he believes$_2$ that she intends$_3$ that he believes$_4$ that the berries are edible*.
> (Scott-Phillips, 2015: 65-66, based on Sperber, 2000, and Grice, 1982; italics and subscript numbering in original)

Scenario five embodies a significant leap according to Scott-Phillips: here Mary has reason to change her behaviour from regular, unremarkable picking to any degree of slightly stylized or exaggerated picking. She now has two intentions, the informative intention (labelled $_3$) she had earlier and the communicative intention ($_1$) to "signal signalhood", which is new to this scenario. However, only if Peter recognises both intentions, "ostensive-inferential communication proper" has emerged:

> *Scenario six*. As per scenario five, including the fact that Mary picks and eats berries in a particularly stylized, exaggerated manner. Because of this, Peter grasps both of Mary's intentions, informative and communicative, as laid out above. As such, *Peter believes$_1$ that Mary intends$_2$ that he believes$_3$ that she intends$_4$ that he believes$_5$ that the berries are edible*. (idem)

Scott-Phillips states that in a world with only the scenarios one to four, there would be no difference between doing things because you need or want to, and doing things in order to communicate with others, since "nobody would signal signalhood" (2015: 67). Only in scenario six is signalhood signalled *and* recognised. At this stage, a form of *interdepence* between signaller and addressee has emerged which Scott-Phillips considers to be a defining characteristic of human communicative interaction: this only obtains if there is mutual recognition of the communicative intention to exchange a particular informative intention, presupposing four and five orders of intentionality to be handled by the speaker and addressee respectively.

After having laid out this strand of reasoning, Scott-Phillips anticipates three types of critique (2015: 68-75): (i) scenario five and six look complicated and cognitively taxing, while we all know from experience that communicating in this way is not; (ii) experimental evidence suggests that children and patients suffering from autism spectrum disorders cannot reason at higher orders of intentionality, but they certainly can be communicatively competent; and (iii) experimental evidence suggests that the limit of orders of intentionality for normally developed human adults lies around five, suggesting that communicative interaction as such is already at the limit. With respect to (i) he points out that there is no a priori reason to assume that something we experience as simple, is also simple in formal terms. He draws a parallel with vision: a formal model of this skill will clearly not be as straightforward as the act of seeing itself feels to us (see Scott-Phillips, 2015: 10). I agree with this in principle, however, we should of course note that this does not work as an argument the other way around: the alleged discrepancy between how vision feels from experience and how complex it may be formally, does by no means entail that everyday communication, feeling easy, should be complicated in formal terms. Besides that, a reason why Scott-Phillips' parallel might not be a feasible one is that vision, being widely spread throughout nature, and pragmatic competence, being unique to humans, require explanations on very different evolutionary time scales. Without a priori excluding anything in the case of human evolution since the divide from the other great-ape lineages, the shortage of evolutionary time is an argument for looking at the most

economical hypothesis in terms of cognitive complexity first (see also Tomasello, 1999).

Regarding (ii) and (iii), Scott-Phillips explains that there are in principle two ways out of the seeming contradictions posed by these types of critique: either the analysis he (and Sperber) set out overcomplicates the matter, or the experimental evidence is wrong and dealing with multiple-order intentionality is much easier and less effortful compared to what is generally assumed. Clearly, Scott-Phillips sets out to argue for the latter. I agree at least partly with him on this point, and do also think that there are issues with the ways in which the experimental evidence has been produced and interpreted (Chapter 6 will deal with this in more detail). However, the two ways out of the seeming contradictions he suggests are not mutually exclusive: besides agreeing that there are some issues with the experimental evidence, I still think that his analysis overcomplicates the matter—in the next subsection I will explain why.

In short, his argument is thus that he sees no possibility to leave out any of the steps of recursive mindreading leading to the emergence of "ostensive-inferential communication proper", as cited above. Therefore, he states, experimental evidence must be at least partly wrong when suggesting that mindreading involving five orders of intentionality is highly cognitively demanding (see Chapter 1, Section 1.2.1), developing late in childhood or even adolescence (idem), and not available to people suffering from certain cognitive disorders (idem). He suggests that ever since mindreading has been investigated experimentally, starting from the late 1970s, the *actual* abilities of human test subjects have been masked by methodological shortcomings. For example, as soon as false-belief tests were carried out "implicitly", i.e. not using explicit questions of the type "Where does she think the sweets are hidden?", the age at which children were able to pass them could be brought down dramatically (from around 3-4 years of age to 12-18 months; see Baillargeon et al., 2010, but see also Apperly, 2011: 29-30 and Heyes, 2012). Similar arguments can be made for people suffering from several psychopathological conditions: different tests have led to better results.

These are indeed important points, which should be kept in mind for the next chapter. However, Scott-Phillips takes them too far in my view: in Chapter 6 I will analyse an "implicit" version of the mentalising test designed under his

supervision, which allegedly demonstrates that normally developed human adults are capable of handling up to eight or nine orders of intentionality effortlessly. His line of reasoning is the same here: he suggests that previous versions of the test have masked the actual performance, and that this is the first one being ecologically valid, thus providing insight into the capacity as it "really" is. He concludes: "There are good reasons, both theoretical and empirical, to conclude that recursive mindreading is not cognitively demanding. More likely, it is, like simple mindreading, something we do habitually and subconsciously, as part of our everyday, low-level perception of the world around us" (2015: 73). I will get back to this in the next chapter. In what follows here, I will take the other of the two suggested paths: instead of (only) criticising the existing experimental evidence, I will (also) scrutinise Scott-Phillips' theoretical analysis of communicative interaction and argue that it is misguided regarding the amount of mindreading complexity it presumes to be necessary.

5.2.3 Individual versus shared intentionality

Brought back to its core, the point I intend to make here can be summarised as follows: whereas the basic cognitive unit in Scott-Phillips' (and Sperber's) analysis is that of a human individual, I argue that the basic cognitive unit of human communicative interaction should rather be understood as *at least two people sharing a great deal of beliefs and intentions*. As a consequence, all the steps suggested to explain how two individuals reach a state of mutual recognition of communicative and informative intentions are rendered superfluous. In other words, where Scott-Phillips sees communicative interaction as a process in which signaller and addressee have separate sets of intentional states which they eventually seek to "pair", I suggest to see communicative interaction as a process in which interlocutors sharing a set of intentional states negotiate what is and what is not part of their shared intentionality.

My view relies for an important part on Clark's work on *common ground* (1996) and *joint projects* (2006), and on Verhagen (2015), who brings together insights from Tomasello's and Rakoczy's notion of *self-other equivalence* (2003; see also Tomasello, 2008; 2014), Searle's *we-intentionality* (1995), and Hutchins'

work on *group-level cognition* (1995; 2006). In the rest of this subsection I will provide some more details on the idea of shared intentionality. After that, I will explore what a communication model based on joint intentionality demands from interlocutors on a cognitive level, especially with respect to mindreading.

The core idea of shared intentionality underlying communicative interaction is that interlocutors consider a particular set of beliefs and intentions to be mutually shared or "common ground". Or perhaps rather: they *a priori* act as if these intentional states are shared, until they have evidence to the contrary. Which set of beliefs and intentions is considered common ground depends on the identity and situatedness of the interaction partner(s). As suggested by Clark and Marshall (1981) there are multiple types of "sources" of common ground. First of all, people can be in the same here-and-now, which is in linguistics generally referred to as sharing the same "Ground" (Langacker, 1990; Verhagen, 2005). In that case they can, for example, use deictic expressions (including pointing and eye gaze) under the assumption that the other can figure out what they mean: they *both believe* that "now" stands for the same moment in time, "here" for the same place, "the book over there" refers to a particular book of which they *both believe* it is that one rather than another, and so on. Another source can be one's personal relationship to someone, formed by a shared history of previous interactions. When speaking to a friend I can, for example, refer to a mutual acquaintance by just using her first name "Susanna", given that we are used to referring to her in this way—in other words, we *both know* who we mean. However, if I want to refer to the same person when speaking to my mother, I may have to say "Susanna Smith", since when my mother and I say "Susanna" we usually mean a different Susanna. In the case of strangers who do belong to the same cultural-linguistic group as we do, most of the common ground is less specific, but we can still assume that a great deal of beliefs are shared, most notably of course the conventional rules of our language. If I produce the sound "huis" (meaning "house" in Dutch) in front of any stranger in the Netherlands, there is a big chance that both our individual histories have assured that we *both think* of a fairly similar concept. In fact, within the Netherlands it is safe to assume this until encountering alongword

evidence to the contrary.[73] Finally, our common ground with all other humans who are strangers and not members of our cultural-linguistic community is still sufficient for some form of shared intentionality to support communicative interaction: this is what we rely on, for example, when looking upward to refer to the sky, or when referring to "food" or "eating" by miming that we take a bite.

Note that people who are in one's personal social network, often also share membership of a cultural-linguistic community, and that both people in one's network and strangers who are members of this community are humans. Therefore, my version of the sources for common ground, adapted from Clark and Marshall (1981) and Verhagen (2015), can be conceptualised as a series of concentric circles:

[73] Verhagen (2005) makes a categorical distinction at this point: (i) all linguistic signals, which rely on sounds (or signs, writing, etc.) being conventionally linked to particular functions within a linguistic community; and (ii) linguistic items that, on top of this, rely on particular knowledge shared between communicators, such as deictic expressions presupposing shared knowledge of the communicative situation or Ground. Tomasello and Rakoczy (2003) rather seem to suggest a continuum where shared knowledge can be very general within a linguistic community on the one side, and very specific between two interlocutors on the other side. This is what I suggest here too.

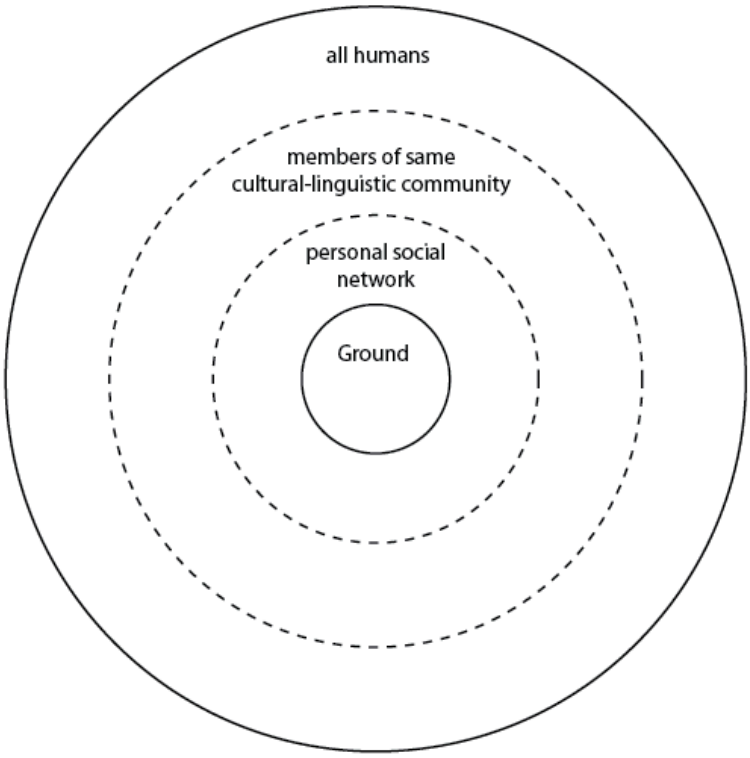


Figure 1 – Types of "sources of evidence" for common ground, adapted from Clark and Marshall (1981) and Verhagen (2015) (I have added the outer circle, slightly altered the categories, and introduced dashed lines for the second and third circles). Which knowledge is considered to be part of the common ground depends on who the interaction partners are. Are they part of the same interaction event, and do they thus share the same here-and-now, i.e. same Ground? Are they people with whom I have a history of interaction? Are they members of the same linguistic and cultural community? Are they humans (or perhaps the question should be: are they normally developed human adults)? The dashed circles indicate that it is possible that interlocutors sharing the Ground can but need not be part of each other's personal social network, and can but need not be members of the same linguistic community.

Groupings like "all Dutchmen" or "all Italians" may yield strong prototypical examples of cultural-linguistic communities, but the definition of such communities also extends to, say, Londoners, dentists, fans of The Police, Oxford students, cricket players, generative linguists, Jehovah's Witnesses, and so on. Whenever a Dutch dentist meets an American dentist, there will be particular knowledge they can consider to be shared on the basis of the community they take part in by virtue of their profession. This probably includes particular experiences and practices, but may also involve a specific

lexicon (sometimes also known as *jargon*, e.g. "endodontics" for root canal therapy) or certain behavioural conventions (e.g. never provide details about a patient's identity). Clark (1996; 2006a) points out that some communities are nested (e.g. Londoners, Brits, speakers of English) and others are cross-cutting (e.g. Oxford students, Police fans, speakers of English). Viewed this way, people are members of many different communities. When two people interact, they generally have a gradation of common ground, based on the amount of shared community memberships: for example, any two Oxford students can assume that the other knows what to wear when sitting exams, whereas Oxford students who are also members of a particular college can not only assume that the other knows about exam dress codes, but also about who used to live upstairs from the old kitchens.

Note that if shared knowledge is indeed a crucial basis for communication, one would expect that people interacting for the first time try to assess whether they share membership of one or more communities, potentially providing them with some common ground. According to Clark (2006a) this is indeed why most conversations with strangers begin by exchanging information about residences, interests, occupations, and so on. At the same time, accents, dressing style, or other aspects of people's overall "habitus", may work as cues (overtly or in disguise) for membership of particular communities. This is not just a matter of finding "something to talk about", but goes much deeper: it is about finding out what the conventional rules underlying the interaction are. To start with, common ground includes knowledge of what to consider as a meaningful signal. Cricket players may draw crucial inferences from gestures hardly even noticeable by outsiders. Or what is just a plate with some used cutlery on it for a member of one community, may to members of another community signal "I haven't finished eating yet". Broadly speaking, these examples are not very different from the fact that speakers of any spoken language recognise speech sounds as

meaningful signals, but only some also recognise "clicks" as such.[74] In all cases, membership of a particular community has, over time, assured that individuals take a particular bit of behaviour as being meaningful in a communicative setting.

Among members of these communities there is thus no need to negotiate these behaviours *qua signal*, to "signal signalhood" in Scott-Phillips' terms. Once they have identified an interaction partner as a member of the same community, hence established a basis for common ground, they can use a signal in the same way as this signal has been used towards them by members of this community. This is what Tomasello and Rakoczy (2003) have referred to as *self-other equivalence*, leading not only to community-wide consensus over what does and does not count as a meaningful signal, but assuring also that linguistic form/meaning pairs become intersubjectively shared within cultural-linguistic communities. In other words, if the principle of self-other equivalence is systematically adopted by members of a community towards other members of this community, this yields a mechanism through which conventions spread reliably. These conventions can be taken in the broadest sense, ranging from what to regard as a signal in the first place, or what to communicate about and what to leave implicit, to more specific conventional associations between forms and meanings such as the gesture "thumb up" signalling a positive attitude, the sound "bal" referring to a round object suited for playing particular games, or the word order "John hits Peter" meaning that Peter is at the receiving end of the action.

The crux in the case of cultural-linguistic communities is that no previous contact between two individuals within a community is needed for them to have a similar set of knowledge states "installed" on their individual cognitive systems. On top of or besides such communal common ground, personal interaction is another way in which shared knowledge can be built, updated, and extended. As soon as two people start interacting, they not only do this

[74] E.g. the Khoeid languages spoken by hunter-gatherers peoples in Namibia (see Voßen, 1997). The difference is of course that such sounds are elements constituting symbols through combination—a very powerful trait of human language—whereas cutlery arranged on a plate is a symbol by itself. There is clearly a lot more to say about how combinations of sounds become conventionally associated with particular meanings, both developmentally and evolutionarily, but that is not relevant for my purposes here.

"against the common ground they believe they already share [but also] as a way of adding to that common ground" (Clark, 2006a: 107, referring to Stalnaker, 1978). This can involve superficial updates (e.g. finding out about a mutual acquaintanceship with "Susanna Smith", enabling unique reference using her name), but also go as far as two people (or a small clique of a few people), who interact frequently, developing their own words, constructions, accent, behavioural rules, and so on.

In this way, we have specific common ground with all people in our personal social networks, besides or on top of the common ground we might have with them by virtue of shared membership of various cultural-linguistic communities, sharing the same "here-and-now" of the interaction ("Ground"), and/or, in the minimal case, being human. An important observation can now be made: once the degree and nature of common ground with a particular interaction partner have been established (possibly through visible cues, accent, conversations about interests, occupation, residence, etcetera), it is possible to derive inferences about this interaction partner's *individual* knowledge, *if need be*. For me as a speaker of Dutch it is possible infer about another speaker of Dutch that she will *know* that the sound "huis" can be used as a signal to draw the attention to some house. Also, I can draw the inference that she will *know* that I *know* this, given that she knows that I am a speaker of Dutch too. Theoretically, I can draw infinitely many inferences like this about what other speakers of Dutch know, what they know that I know, what they know that I know that they know, and so on (this point is also made in different forms by Clark, 2006a, and Verhagen, 2015, both referring to Lewis, 1969). However, this is not what I need to do *by default* before communicating with them, given that "as such" these inferences add nothing new: each of them is a *derivate from* the already existing common ground between all speakers of Dutch, not a step *towards* the emergence of such common ground. The same holds true for any form of common ground. When I sit behind my desk and my office mate has gone out, before going home I can leave a note on his desk saying that I won't be "in HQ" before next Tuesday, *knowing* that he will *know* that I *mean* our office by HQ ("Head Quarters"). And he will know that I know that he knows what I mean. Also, when we both sit behind our desks, I can point towards the windowsill, where our coffee machine is situated. He may

nod, which I can take as an indication that he indeed would fancy a cup. This works because I *know* that he *knows* where our coffee machine stands, and I *know* that he *knows* that I *know* this, and I *know* that he is a coffee drinker, and that he *knows* that I *know* that he is. All these inferences about who knows what about the meaning of "HQ", or the location of the coffee machine and the desire of it being put to use, can be *derived from* our personal common ground and the common ground provided by our co-presence in the same here-and-now. Most of the time, we never get around to drawing such inferences, although it is possible to think of contexts in which we might do so. For example, imagine him saying "no thanks" once I put the freshly brewed cup of coffee in front of him. We might enter a conversation about what went wrong in the previous communication: I could say that I *thought* he *wanted* coffee because he nodded when I pointed at the coffee machine, to which he might reply that he *understood* I *wanted* to lower the blinds and therefore pointed at the window. In this example, intentional reasoning seems to enter the stage only because of a misunderstanding inciting some reflection on differences in perspectives. Towards the end of this chapter I will follow up on this point of explicitly constructing the different perspectives involved in an interaction event, also in the light of the distinction between holistic and compositional complexity introduced in the previous chapter.

The analysis at this point closely resembles what I have covered in Chapter 1 by introducing the "Schelling mirror world" (following Levinson, 2006). Interlocutors toss into the Schelling mirror world a piece of behaviour (words, gestures, facial expressions, or otherwise) which they consider suitable for letting the other draw the desired inferences. "Meetings of the mind" (Levinson, 2006: 49) in this world can, as discussed, occur by virtue of having a shared sense of salience. We can now see that such a sense can be derived from common ground using the principle of self-other equivalence: I can pick the behaviour *I* find most appropriate in order to evoke a particular inference in my interlocutor, which is, given our common ground, by definition also the most appropriate inference in her eyes. Only if this goes wrong, may we need to figure out what happened asking ourselves what the other was thinking, and possibly what the other thought we were thinking, what the other thought we were thinking the other was thinking, and so on. Even without a previous

misunderstanding we may want to reflect on the communicative situation in such a way, perhaps for purposes of analysis or strategic planning ahead. Crucial, however, is that complex mindreading or intentional reasoning comes with such reflection, and is not relied on by default.[75]

Or is it after all? How often do we need such reflection? Are we not in need of reflection on the communicative situation all the time, either implicitly or explicitly? As also mentioned in Chapter 1, various mechanisms seem to be saving communicators from having to engage in cognitively demanding mindreading most of the time. First, following the idea of *relevance* as worked out by Sperber and Wilson (1995; 2002) most communication comes down to the signaller picking the first (i.e. the most relevant) expression that comes to mind and the addressee picking the first interpretation that comes to mind. Sperber and Wilson themselves argue that sophisticated mindreading skills are needed for this process. However, I agree with Apperly that they seem to overlook that especially their updated account of relevance (2002) renders mindreading almost entirely superfluous: given that interlocutors are "in complementary predicaments", it is "a good bet for [them] to follow their own individual cognitive paths of least resistance" (Apperly, 2011: 115). Both pick the most relevant expression or interpretation first; if this does not lead to a satisfactory result, they can try the second-most relevant expression or interpretation, thus working downwards on the gradient of relevance. This fits with what various psycholinguists studying "alignment" have found: Pickering and Garrod (2004) argue explicitly that due to these mechanisms (relevance, alignment) interlocutors can refrain from constantly making inferences about the other's mindstates (see also Apperly, 2011: 116). Besides this, Apperly makes another point that is relevant here: in everyday interaction, we do not have to go to the bottom of everything. Rather, we seem to work with representations

[75] Note that Tomasello seems to come to a similar conclusion in his 2014 book (see especially page 38). However, it is unclear from this passage, and from the parts of his 2008 book that he refers to here, what his exact position is in "diachronic" and "synchronic" terms. Does he see shared intentionality as a feature that emerges from and is conceptually underpinned by layers of embedded intentionality, but is in practice usually not decomposed into these constituting layers? In that case he would defend the same diachronic story as Scott-Phillips (2015) does, but a (somewhat) different synchronic one. Alternatively, his view could be that both the diachronic and synchronic stories can do without these layers, which is what I suggest in this chapter and in the Conclusion below.

that are "good enough" for the interaction to keep going, but no better (2011: 114-119 and personal communication). If required in a conversation, interlocutors can together work out a particular point in more detail, aiding and steering each other in the desired direction turn by turn. In Levinson's (2006) terms: many conversations do not have a "signal-response" structure, but rather one of "testing-adjusting-retesting".[76]

What is more, not only are signaller and addressee experienced in choosing the most relevant cues and interpretations, the linguistic tools they have available also contain a wealth of such accumulated "experience". After all, they have emerged in the course of generations of language users attempting to coordinate the perspectives of themselves, their interlocutors and possibly third-party referents. Verhagen (2005; 2015) shows, for instance, for deixis, negation, and particular discourse connectives how they work "argumentatively" in the process of negotiating how (potential) deviations from the common ground can be resolved, in order for the interaction to be able to continue. This, then, introduces the issue central in the second part of this chapter: starting from a set of shared intentional states as defined by the interlocutors' common ground, it is possible to single out and highlight differences between individuals and the non-shared part of each of their intentional states, thereby enabling negotiation about how the common ground should be updated. This is the domain of viewpoint coordination in discourse, for which language has a great number of specific tools, some of which have been discussed in the previous chapters. The next section discusses a conceptual model for analysing this class of linguistic tools.

The final question remaining for this subsection is what kind of structure is required on the cognitive level for this alternative model of interaction, starting from shared intentionality or common ground, to work. It is important to note that I am not suggesting that processing efforts needed to determine what is and what is not part of the common ground with an interaction partner

[76] This again fits well with Clark's work on conversations as *joint projects*, in which interlocutors implicitly commit themselves to particular goals (which could be anything from setting a coffee meeting for tomorrow to cooperatively completing a complex building task) and converse about how the common ground has to be updated in order to achieve these goals (see Clark, 2006).

are always little or insignificant compared to those needed for ostensive-inferential communication according to Scott-Phillips. However, I argue that such efforts start at or close to zero by default (given that common ground is assumed a priori), and are then scaled up if necessary. In contrast to this, Scott-Phillips suggests that we use "recursive mindreading up to level five" by default. At least regarding the "synchronic" part of the story (see Section 5.1 above), a list of requirements needed for my alternative model would look like this:

(i) quick and fairly accurate abilities of distinguishing between individuals belonging to our own social network and/or particular cultural-linguistic communities we participate in;

(ii) abilities to keep track of former interactions we had with others, and access these "records" during interaction;

(iii) the capability to apply the principle of self-other equivalence, as needed to operate in the "Schelling mirror world";

(iv) the ability to reflect on individual perspectives and how they deviate from the common ground when prompted to do so, especially in the light of misunderstandings or other special circumstances that require scaling up of processing efforts.

In the conclusion of this chapter I will get back to this list from a "diachronic" perspective, thus exploring consequences for our evolutionary story.

With all of this in mind, let us now briefly return to Scott-Phillips' and Sperber's example of Mary and Peter communicating about berries being edible. I suggest that Mary and Peter live in a world where particular beliefs and intentions can be considered to be shared by default. Therefore, they can under normal circumstances bypass any intentional reasoning about what the other believes, what the other believes that they believe, and so on, including what the other will believe to be a meaningful signal and what not. If Mary wants Peter to believe that the berries are edible, she can just carry on picking and eating. Only if she wants to *deviate* from this default do things become more

complicated: for example, if she wants to mislead him about the berries being edible, or if she realises that he should not be thinking they are.[77] In this case, she can reflect on the situation, deriving from the common ground what Peter will believe and reasoning about how this will change according to her behaviour. In other words, she can begin negotiations with Peter about how the common ground should be updated without further difficulties, and only if need be will she unpack the default situation, thereby scaling up processing effort.

5.3 Coordinating mindstates in discourse

In this section I will propose a conceptual model for analysing the broad and diverse class of linguistic items capable of viewpoint coordination. The purpose of my model is not primarily to introduce another practice of drawing schemas, next to for example Dancygier's (2012) narrative spaces framework (as used in Chapter 3 and 4) or Fauconnier's (1985) mental spaces. It *can* be used for schematically representing individual viewpoint configurations as prompted by particular linguistic items, and I consider it illuminating to do this a few times when explaining its details, but the main purpose of including it here is to make a point about the general structure and working of linguistic interaction. This point is, to provide an anticipatory summary, that many linguistic items not only work to draw attention to some object or concept in the world—a function often described as *reference*—but also, and mostly at the same time, to provide and manage *perspectives on* or *stances towards* these objects or concepts. In Chapter 3 and 4 I have referred to this as the polyphonic nature of discourse. As such, the point that most language usage also entails viewpoint coordination is recognised by many linguists and narratologists; however, focus has mostly been *either* on how signallers and addressees mutually coordinate their perspectives (see e.g. Langacker, 1990; Sweetser, 1990; Verhagen, 2005), *or* on

[77] In his original discussion of the example, Sperber (2000) does address the special case of misleading someone else. However, this does not alter his analysis that even in basic, straightforward cases intentional reasoning up to five orders is needed.

how third-party perspectives are represented (e.g. Fludernik, 1993; Bal, 2009; Hühn, Schmid and Schönert, 2009).[78] The model of linguistic interaction proposed here integrates these functions all at once, instead of approaching them as distinct phenomena, thereby capturing the polyphonic nature of discourse in a conceptual model—giving it a schematic "face", as it were.

5.3.1 Dyadic and triadic communication

Communication in non-human animals typically involves a sender producing some observable behaviour (the "signal") that increases the likelihood of a receiver responding, i.e. behaving, in some particular way—for example, a bird signalling to a competitor to stay away from his territory. When the benefits of such a pattern of linked behaviours outweigh the costs for both senders and receivers, a (relatively) stable communication system may emerge. Thus, most non-human communication is about "regulating and assessing the behavior of others" (Owings & Morton, 1998: i). At this very basic level, the conceptual space needed to characterise communication is one-dimensional: no other dimension than that of the sender-receiver relationship is necessarily relevant to characterise a signal and its causes and effects. In the words of Tomasello (2008: 23), animal communication is mostly "dyadic": by far the majority of cases can be explained in terms of regulating others' behaviours without having to take into account attention (let alone *joint attention*) to any objects of reference.

[78] An exception is Dancygier (e.g. 2012), and to some extent also Vandelanotte (e.g. 2009): their approaches also integrate insights from linguists' interaction models with narratological views on speech and thought representation.

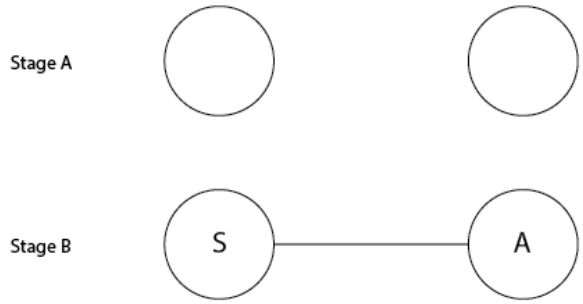


Figure 2 – Schematic depiction of "dyadic" communication as found in non-human animals. In Stage A there is no communication between the two depicted subjects (circles). Stage B shows the situation in which the sender (S) transmits a signal to the addressee (A) in order to influence the latter's behaviour, thus establishing one-dimensional communication. (Possibly, this induces a response signal: the subject on the right then becomes S and the subject on the left A.)

In contrast, human communication is prototypically "triadic" (Tomasello, 2008: 23) as it is by default *about* referents in the (shared) world outside of the communicators and their interaction. Following this idea, the conceptual framework needed to characterise normal human communication should thus at least be two-dimensional: apart from the relation between the communicators, the relation to the world must be taken into account to characterise signals and their causal connections. In other words, at the heart of interaction using language lies joint attention to some object of conceptualisation (person, event, relationship, etc.) and negotiating a particular stance towards this object.[79]

[79] Nonetheless, there are some instances of communication in non-human animals where functional reference to objects in the "outside world" does seem to play a role; a well known case is that of vervet monkey calls referring to different types of predators (see Seyfarth, Cheney, and Marler, 1980; Dennett, 1987: chapter 6). Conversely, humans also regularly engage in purely dyadic communication, such as greetings ("Hi!") or warnings ("Watch out!"). However, as Owings and Morton (1998: 211) argue, functional reference in animal communication should not be analysed as providing information about entities in the world, since it would confuse short term with long term causation; objects such as a snake in a snake alarm call should be seen as "long-term validators of the signal's utility", not as real-world correlates of signals which are causally involved in the receiver's response to the signal. In human communication, however, triadic communication *does* prototypically involve real-world objects of joint attention.

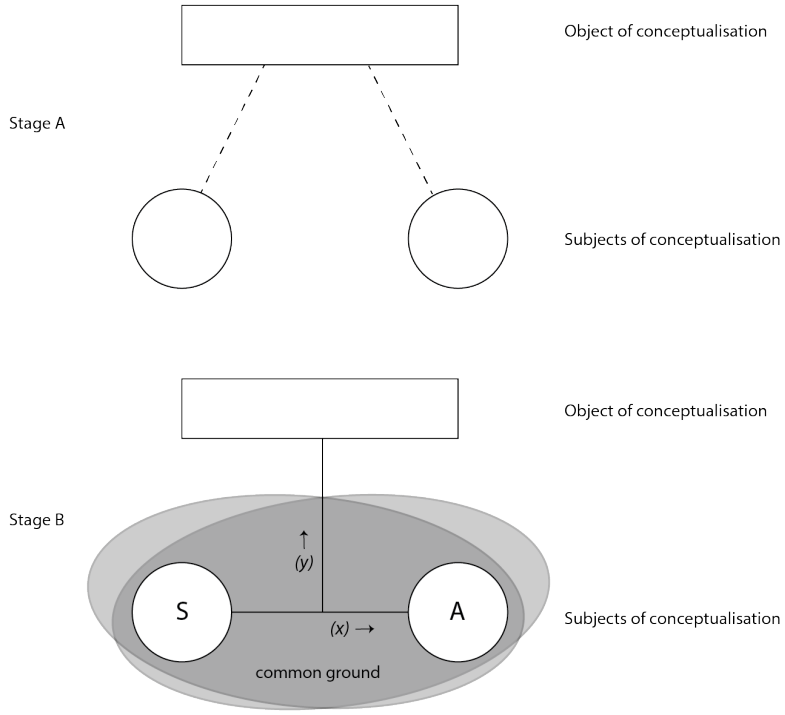


Figure 3 – In Stage A, where no communication takes place, the two subjects (circles) both have their own views (dashed lines) on some object (rectangle). Stage B depicts triadic communication: the signaller/speaker (S) and addressee (A) both assume a set of shared beliefs (the overlapping part of which is the *common ground* discussed in Section 5.2 above) and subsequently negotiate how the common ground should be updated with respect to the object. As will be discussed below, the used signals typically reflect both aspects of and operations on the relationship between S and A (the *(x)*-axis) and on the relationship between the common ground and the object (the *(y)*-axis).

In simple terms, the two-dimensional conceptual space depicted in Figure 3 makes a distinction between the "intentional" aspect of language, its capacity to be *about* some object in the world, and the "(inter)subjective" aspect of language, according to which sender and addressee negotiate a particular stance *towards* this object. These aspects are depicted by the *(y)*- and *(x)*-axes respectively.

The field of cognitive semantics also embraces the idea that the proper characterisation of language use requires a two-dimensional conceptual framework, but its background and history differ somewhat from the biological and psychological considerations made here. In fact, the cognitive view was explicitly designed by Langacker (1987) in opposition to so-called "objectivist"

approaches to semantics, which held that meaning in natural language could be fully characterised in terms of no more than its relation to the/some world (its contribution to "truth conditions"). Objectivist semanticists were thus precisely ignoring the "perspectival", "(inter)subjective" *(y)*-axis. This axis is indispensable in Langacker's view, since he claims that different "perspectival construals" are just as inherent components of linguistic meaning as reference is.[80]

What I will argue now is that a proper characterisation of viewpoint management in discourse (and of linguistic elements supporting viewpoint management), requires recognising a third dimension. I will begin with a single case that presents a problem for the two-dimensional model, and show how the addition of a separate dimension relating the present communicative situation to other ones, provides a straightforward solution. Subsequently, I will show that this new model also provides a very natural framework for the analysis of other items and viewpoint configurations.

5.3.2 Speaker commitment and viewpoint embedding: Dalabon and English

Consider the following utterance in the Australian language Dalabon and its English translation as suggested by Nicholas Evans (class lectures 2009, brackets in original):[81]

[80] Objectivist approaches to language thus in fact also assume a one-dimensional conceptual framework for the analysis of meaning in natural language, but highlight the *other* dimension (i.e. the *(y)*-axis) instead of the dimension I have suggested to be relevant for non-human communication (the *(x)*-axis). Verhagen (2005) extended Langacker's model by including a systematic distinction between the viewpoints of the sender and addressee, in order to bring out the fact that construal is not (just) a matter of a single viewpoint ("subjectivity") with respect to some object, but one of mental coordination between signaller and addressee with respect to an object of joint attention ("intersubjectivity"). In hindsight, we can say that the framework proposed by Verhagen (2005: 7) represents a merger of the biological and cognitive-semantic views of human communication.

[81] I thank Nicholas Evans for permission to use this example in this context. See Evans (2010) for more examples of elements for viewpoint coordination (esp. ch. 4), and for the glossing method used.

(1)	Ka-h-kangurdinjirrmi-nj	yangdjehneng	bûrra-h-marnû-dulu-djirdm-ey
	3SG-ASS-GET.ANGRY-PSTPERF	SUSPENDED-COMMITMENT	3DUHARM.SUBJ>3SGOBJ-ASS-BEN-SONG-STEAL-PSTPERF
	He got upset [because]	*[he thought that]*	*the two of them had stolen his song*

(2) He got upset [because] [he thought that] the two of them had stolen his song.

"Because" and "he thought that" are inserted in the paraphrase by Evans. I will briefly discuss the causal marker "because" in note 82, but focus on the insertion of "he thought that" in detail first. It is clear that the lexical unit *yangdjehneng*, glossed as "SuspendedCommitment" does not literally mean "he thought that", but rather conveys the message: "I, speaker, am not committed (to what I am going to say now)". A paraphrase closer to the original expression is thus:

(3) He got upset [because] [I, speaker, am not committed to this:] the two of them had stolen his song.

At first sight, it may seem remarkable that Evans renders the lexical unit that functions as a marker of suspended commitment with a complementation construction in English—are the two indeed equivalents? To illustrate that, in an important sense, they are, consider the differences between the more idiomatic translation in (2) and the more literal one in (3). The absence of "he thought that" in (3) does not mean that "he" no longer had the thought that "the two of them had stolen his song". In fact, awareness of the information in the second clause is equally implied in (2) and (3); if "he" had not had that thought, the stealing of the song could not have caused him to be upset. In both the Dalabon and English versions the speaker invites the addressee to view the information about "the two of them" having stolen the song from the perspective of a third party, namely "he" introduced at the beginning of the sentence. However, there are differences in the degree to which this is accentuated: the coordination of a third-party perspective is significantly more

pronounced in the idiomatic English translation in (2) compared to the Dalabon original (1) and its paraphrase in (3).

In a similar sense, the Dalabon element *yangdjehneng*, the English phrase "I, speaker, am not committed to this", and the idiomatic pattern of sentence complementation using the stance verb *to think* all three negotiate a(n epistemic) stance of the speaker towards parts of the presented content: in (1), (2), and (3), the speaker does not assert as true that "they had stolen the song". Yet the difference is again in the accentuation: in (1) and (3) the tempering of commitment by the speaker is realised "on stage", whereas in (2) this remains implicit.

In both the Dalabon and English versions the speaker thus invites the addressee to view the information about the two of them having stolen the song from the perspective of a third party, namely "he" introduced at the beginning of the sentence. However, there are differences in the degree to which this is *accentuated* or *profiled*: the embedding of the information in a third-party perspective is done explicitly by means of a particular syntactic construction (complementation) and a particular matrix predicate (*thought*) in English, with the speaker's reduced commitment remaining more implicit, while the latter is precisely being profiled by the Dalabon element *yangdjehneng*, with the third party's relatively higher degree of responsibility remaining more implicit. In short, these conventional ways, in these two languages, of distributing responsibility for a piece of information over the speaker and another party are each other's mirror image: what is explicitly "put on stage" and "what is left to inference" is so to speak reversed. But the totality of what is communicated with

these structurally very different expressions, is very much the same, in particular the connections between different relevant viewpoints.[82]

Can both the similarities and the differences between these expressions be stated in a single analytic framework? If we try to do so using the two-dimensional model of triadic communication in Figure 3, it soon becomes clear that this requirement cannot easily be satisfied. If the function of the element *yangdjehneng* ("I am not committed to this") is straightforwardly characterised as the speaker signalling to the addressee (*(x)*-axis) what his stance is towards (*(y)*-axis) the object of conceptualisation (i.e. "the two of them had stolen his song"), the associated heightened responsibility for this view of the third party ("he", the one who got upset), is necessarily left out. The reason is that this third party is only present in this model as an element of the situation being talked about, as an *object* of conceptualisation, and not as another *subject* taking a view on this situation.

Conversely, the representation of the complementation construction in (2) ("he thought that...") does not work very well in a two-dimensional framework either, conceptually. In (2) "he" is clearly not only an object of conceptualisation (we are presented with *what he thought*), but at the same time, "he" should not be seen as a subject of conceptualisation in the same sense as the speaker and addressee either. After all, the *negotiation* of a stance towards the object of conceptualisation takes place between speaker and addressee, meaning that the speaker can, as a part of this negotiation process, *invite* the addressee to consider the perspective of a third party on some aspect of the object of conceptualisation. However, perspective can never shift completely to this third party in the course of the modelled interaction event (cf. the way it can jump

[82] The phenomenon of marking explicitly only some aspects of what is to be conveyed is, of course, not limited to viewpoint expressions; on the contrary, it is quite general and well-documented for various conceptual relationships, including causality (see e.g. Verhagen, 2005). It should be noted that an analysis similar to the one given above applies to the pair (i) *He got upset; the two of them had stolen his song* and (ii) *He got upset because the two of them had stolen his song*. The conceptual representation of both (i) and (ii) contains a causal relationship (otherwise no coherent interpretation seems possible), but this is only marked explicitly, "on stage", in (ii). The difference between the Dalabon and English idiomatic ways of expressing both viewpoints and causal relations can be characterized as a difference in the available tools, and in the conventional rules for using them in the different languages. See also Wilkins' (1986) discussion of "particle/clitics" for criticism and complaints in Aranda, another Australian language, and his argument that these encapsulate "culture specific modes of thinking" that become clear when their use is explicated.

from one character to another in a novel). In other words, the view of the third party "he" can be instrumental in the speaker's and addressee's negotiation of a stance towards the object of conceptualisation, but "he" is himself not a participant in this negotiation process. All in all, the common problem when representing the sentences (1), (2), and (3) seems to be that in a two-dimensional conceptual model of communicative interaction, third-person conceptualisers can only be situated either at the level of the object of conceptualisation, or at that of the speaker and addressee, while in fact they normally belong to neither.

I therefore propose to treat other subjects of conceptualisation not as additional entities in the two-dimensional space, but as implying the addition of a third dimension, which links third parties exhibiting intentionality towards the relevant object of conceptualisation to the level of the negotiation process between speaker and addressee. The basic idea is captured in Figure 4:

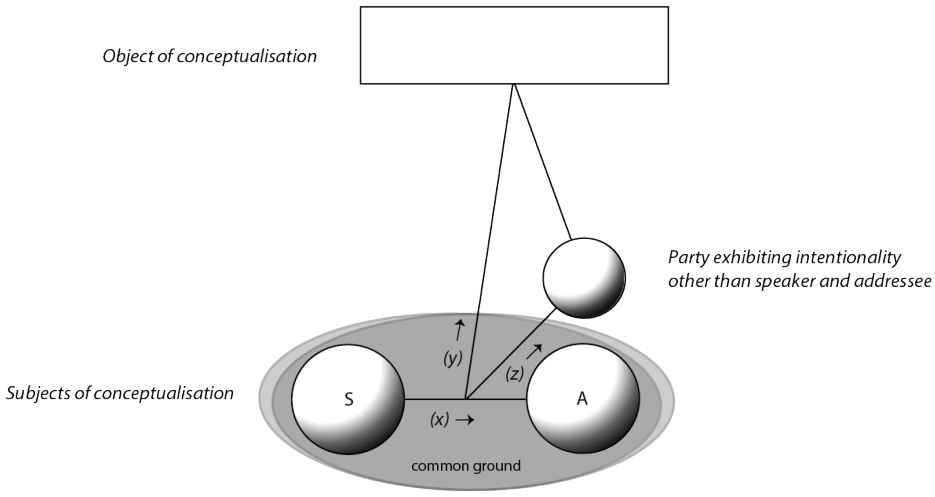


Figure 4 – The three-dimensional conceptual model of interaction featuring a non-speaker, non-addressee subject of conceptualisation.

We conceive of the third person represented in Figure 4 as a subject of conceptualisation in exactly the same way (i.e. with the same cognitive capabilities, including intentional reasoning) as the speaker and addressee. Moreover, the object of conceptualisation for this subject is (at least in part) the same as the one that the speaker is inviting the addressee to consider, capturing

the idea that the speaker presents the situation to the addressee *from a third-party perspective*.

This basic model provides the conceptual space to mark precisely the similarities and differences between the Dalabon and English viewpoint items discussed above, regardless of the fact that they belong to completely different language systems. A graphic representation of these forms can be found in Figure 5 and Figure 6, respectively.

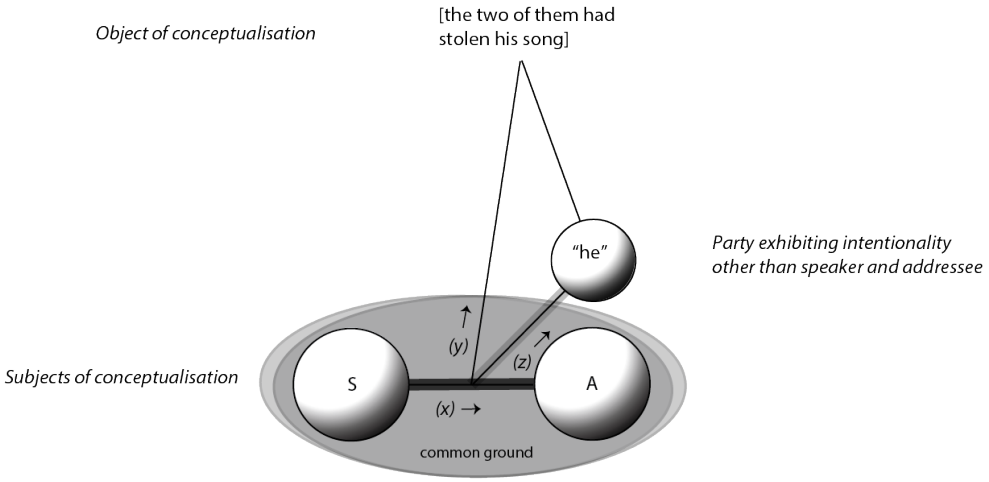


Figure 5 – Dalabon: *yangdjehneng*

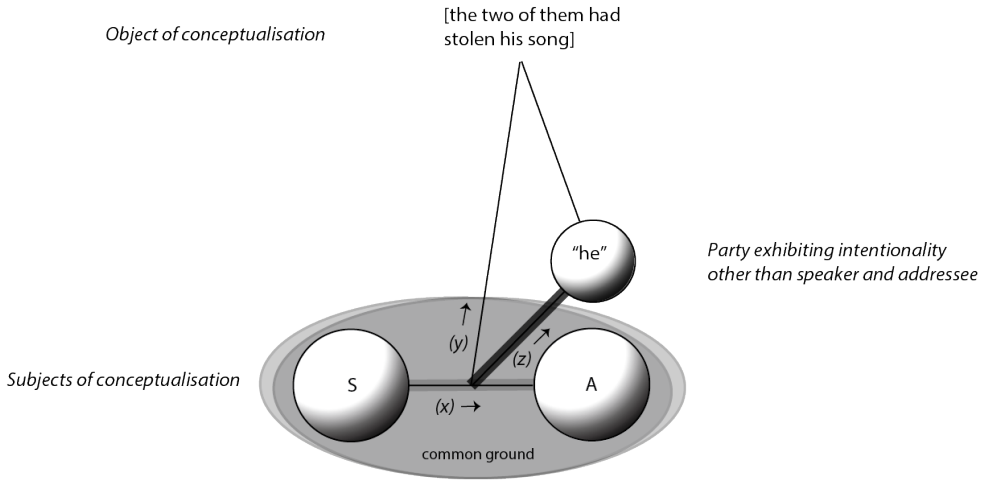


Figure 6 – English: *to think* + complement

As in the two-dimensional model, the negotiation of epistemic stance performed by both the Dalabon and English elements is highlighted along the *(x)*-axis between S(peaker) and A(dressee). In Figure 5, this axis is marked with a dark line, indicating foregrounding of the speaker's epistemic stance by the Dalabon element *yangdjehneng*, glossed as "I, speaker, am not committed". In Figure 6 this axis is marked with a light grey line, indicating that the English complementation construction with *to think* does impact upon the negotiation of epistemic stance between S and A, but in a less pronounced way than the Dalabon element does.

What is new in Figure 4, 5, and 6 compared to the two-dimensional version in Figure 3 is the *(z)*-axis connecting the *(x)*-axis to a third party, in this case the person referred to using "he" and "his" in Evans' translation in (2). On this axis, the reverse pattern obtains with regard to profiling: whereas in Dalabon this third-person perspective is only implied, indicated by a light grey line along the *(z)*-axis in Figure 5, in English it is explicitly realised "on stage", indicated by a dark grey line on the *(z)*-axis in Figure 6. Thus, thanks to the additional *(z)*-axis, we now have a single format for representing that both the Dalabon and the English versions of the utterance invite the addressee to consider the third person's perspective on the matter talked about, i.e. the (actual or imagined) stealing of the song by "the two of them", and that they do so in different ways, by highlighting what parts of the configuration are linguistically marked in each language, and which are implicit, but made inferable.

5.3.3. The general model

When I first introduced the three-dimensional model, I stated (below Figure 4) that the additional intentional party is a subject of conceptualisation whose perspective is instrumental in the speaker's and addressee's negotiation of how the common ground should be updated with respect to an object of joint attention, without himself being a participant in this negotiation process. However, this third party may himself be represented by the speaker as being involved in another communicative interaction event, and in fact, this party

may be talking or thinking about yet another interaction event. Thus, we may in principle expect to encounter more elaborate constellations of several subjects all in some way considering the same object of conceptualisation from different viewpoints, and affecting (more and less mediated through the viewpoints of others) the negotiation between S and A of epistemic stance, attitude, etcetera. Such a constellation is depicted in Figure 7:

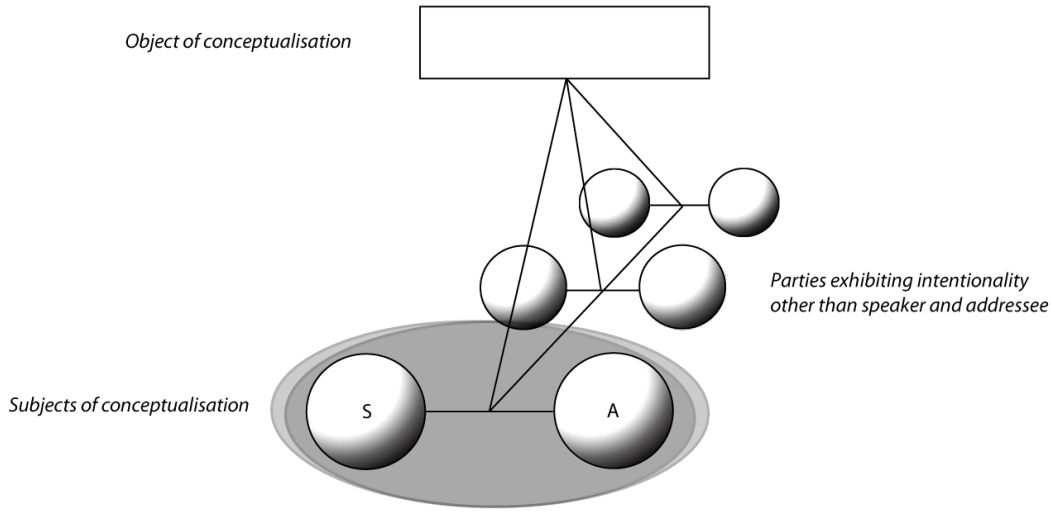


Figure 7 – While communicating about some object of joint attention, S and A may refer to other interaction events, each featuring *their* participants.

In the situation depicted in Figure 7, viewpoints from the other interaction events must, in one way or another, be relevant to how S and A assess their object of conceptualisation. As an example, imagine two people, Simon and Arran, waiting for a man named John to show up at their appointment. Simon has seen John the day before and when the appointment was mentioned, John's daughter Mary kindly reminded her father that he is *always* late. Now Simon says to Arran that "John assured Mary that he would be on time". Figure 8 depicts this situation schematically:

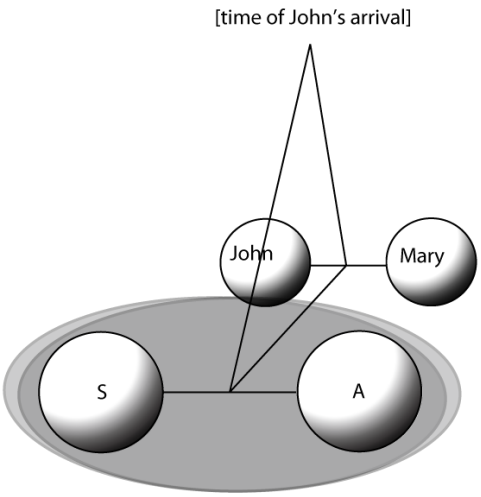


Figure 8 – "John assured Mary that he would be on time"

In this example, there is thus another interaction event being called up within the current interaction: Simon coordinates his perspective on John's expected time of arrival with Arran by referring to how John was coordinating *his* perspective on his arrival time with Mary the day before. He could have done this in numerous alternative ways, for example by saying "John said to Mary: 'I'll be on time'", "When I saw John and Mary, John thought he would be on time", "John will be on time. He promised Mary", and so on. All these alternatives feature a slightly different distribution of responsibility for what John said and the amount of commitment made by Simon to John being on time, given that some elements modify the nature of the relationship between third-person conceptualisers and the speaker and addressee in different ways. Thus, in this example, the use of indirect discourse and the choice, by the speaker, of the verb "assure" (unlikely to have been used by John himself), indicate some degree of co-responsibility of (and interpretation of John's utterance by) the speaker, higher than with the use of a neutral verb of speaking and direct discourse (e.g. "John said to Mary: 'I'll be on time'"). These differences are as such interesting from a semantic, grammatical, or narratological perspective, but go beyond the point I want to make here—which is that all alternatives feature different linguistic elements (words,

grammatical constructions, patterns of speech and thought representation) with different meanings, leading to a variety of overall interpretations, by operating on parameters within the same conceptual space: the relationship between the speaker and addressee (*(x)*-axis), the relationship of the communicative interaction with other interactions featuring third-parties (*(z)*-axis), and all of their perspectives on the Object of conceptualisation (*(y)*-axis).

5.3.3 Thoughtscapes and the model

Some of the examples provided in Chapter 4 were drawn from news sources reporting on the "Pistorius case", the tragic shooting of Reeva Steenkamp by athlete Oscar Pistorius. Recall that the difference between the competing versions of what happened during the night of the shooting completely depended on the construal of Pistorius' intentional state at the moment of pulling the trigger: did he *think* he was shooting at a burglar or did he *know* his girlfriend was behind the bathroom door? The news media not only reported the perspective of the athlete, but also of police detectives, spokespeople, journalists, witnesses, family members, and so on. The result was what I termed a "thoughtscape", a series of perspectives that are mutually connected and embedded in various ways. What could be found in the news reports was what I referred to as "polyphonic" discourse representing this thoughtscape: all kinds of linguistic elements were doing some part of the labour of coordinating the involved perspectives, including grammatical constructions (such as complementation and inquit-constructions), various patterns of reported speech and thought, lexical items (such as *allegedly* and *accidentally*), tense, modality, and more. One of the examples was the following opening sentence from a South-African press release:

> (4) Athlete Oscar Pistorius **allegedly accidentally** shot dead his girlfriend at his house in Pretoria on Thursday morning, *Beeld.com* reported.
> (SAPA, 'Oscar Pistorius shoots girlfriend: report', 14 February 2013)

As a whole, (4) fits a particular embedding pattern, termed an inquit-construction in Chapter 4, in which the reported clause precedes the reporting

clause (underlined). The inquit-construction does part of the viewpoint coordination: it attributes the claim that "Pistorius allegedly accidentally shot dead his girlfriend" to the perspective of newspaper *Beeld.com*. However, there are more viewpoints being coordinated. It is implied by the adverbs *allegedly* and *accidentally* (boldface) that some external source *claims* that Pistorius *did not intend* to shoot his girlfriend. In other words, already on the basis of one sentence, readers are confronted with a thoughtscape involving three viewpoints, without even counting the perspective of the speaker (i.e. the journalist who wrote the sentence).

I will first abstract from the reporting clause of the inquit-construction and concentrate on the reported content:

(5) Pistorius allegedly accidentally shot dead his girlfriend.

The words *allegedly* and *accidentally* are instantiations of what was in Chapter 4 described as viewpoint packages, words implying a topology that introduces one or several extra viewpoint layers. In the case of *accidentally*, it is given in this topology that an agent did not intend X, but it is known that the outcome is X. In actual usage this topology is assimilated (through *blending*; see Chapter 3) with details provided in the immediate context. For example, readers of (5) will blend their knowledge of the topology of *accidentally* with "Athlete Oscar Pistorius" and "shot dead his girlfriend", and take it that he shot her dead, but *did not intend* to do so. In this way, using *accidentally* the speaker invites his addressee to consider the perspective of a third party, in this case Pistorius. Since this is not highlighted explicitly, in the depiction below a light grey line is used along the *(z)*-axis:[83]

[83] The word *accidentally* clearly also negotiates a relationship to an object in the world on the *(y)*-axis, but in my discussion here I will abstract from these relationships and focus on those indicated on the *(x)*- and *(z)*-axes.

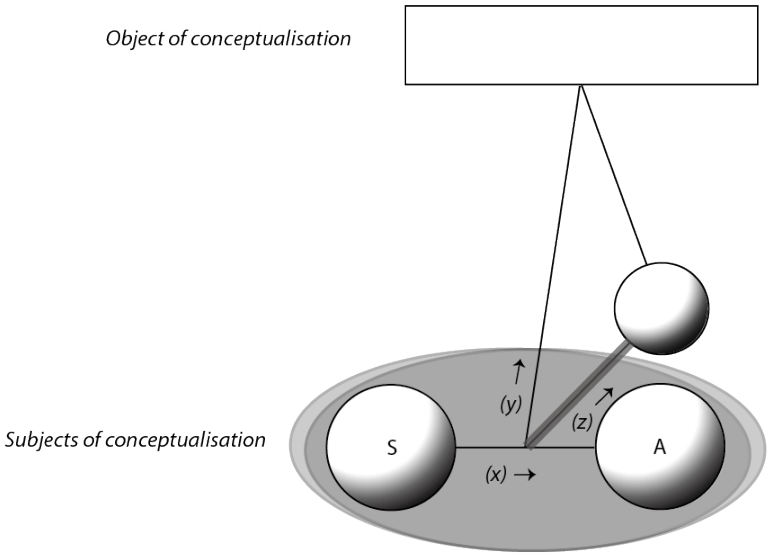


Figure 9 – accidentally

In a similar way, in the topology of *allegedly* it is given that some source X, not the speaker, asserts the content under the scope of this adverb. This topology can be elaborated to various degrees. The identity of source X can be given in the context, or left open, as is the case in (4): readers of this sentence will understand that some external source not specified here claims that Pistorius accidentally shot dead his girlfriend.[84] In that sense *allegedly* shows strong similarities to the Dalabon element *yangdjehneng* cited in (1) above. It suggests the presence of an extra viewpoint, lowers the epistemic commitment the speaker makes to the related content, and, indeed, could also be "translated" using a complementation construction:

(6) It is claimed that Pistorius accidentally shot dead his girlfriend.

[84] Recall that in Chapter 4, Section 4.3.3, an alternative reading of (5) is discussed next to the one given here. However, distinguishing between these two options is not relevant here.

In terms of the present model, *allegedly* is thus a linguistic cue that negotiates a particular epistemic stance of the speaker, while at the same time inviting the addressee to consider the perspective of a third, in this case unspecified, party. It operates along the *(x)*- and *(z)*-axes, albeit without a particular emphasis on either. Consider the schematic depiction in Figure 10:

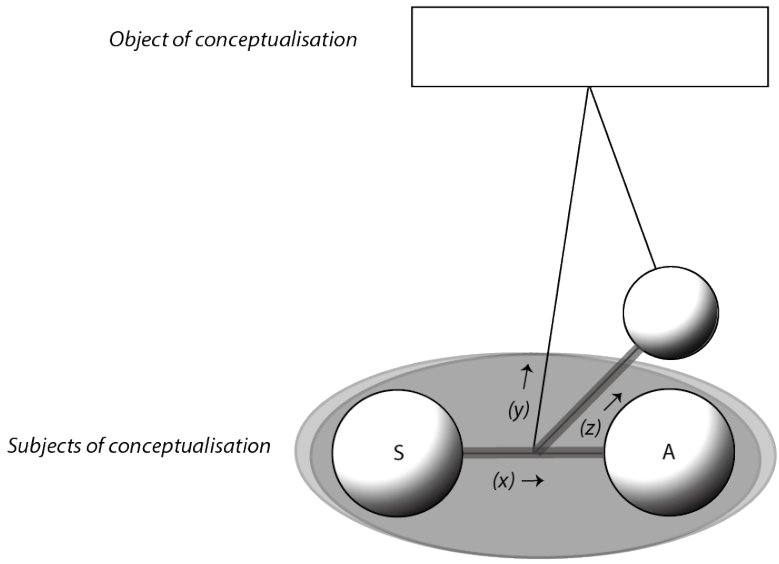


Figure 10 – *allegedly*

A schematic rendering of (5), involving at the same time the viewpoint coordination effected by *accidentally* (i.e. the athlete *not intending* to shoot his girlfriend), is also possible in the proposed conceptual space. This involves the inclusion of one more viewpoint along the *(z)*-axis, which can be done as follows:

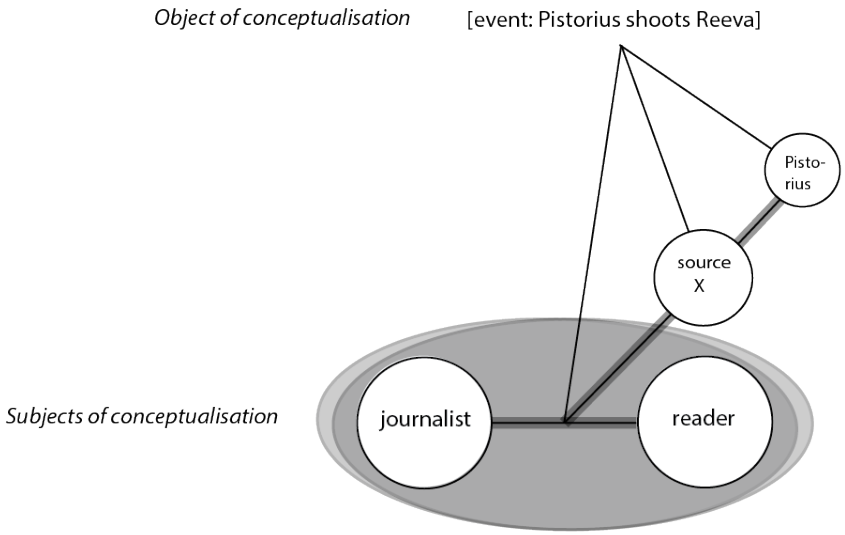


Figure 11 – Sentence (5)

Finally, the perspective of *Beeld.com*, which is coordinated with respect to the reported content using the inquit-construction in (4), can be added to the picture. Given that the introduction of the perspective of *Beeld.com* takes place explicitly, "on stage", a dark grey line is used here along the *(z)*-axis. The introduction of the two additional perspectives (Source X and Pistorius) as well as the negotiation of epistemic commitment is done implicitly, "off stage", hence the light grey lines:

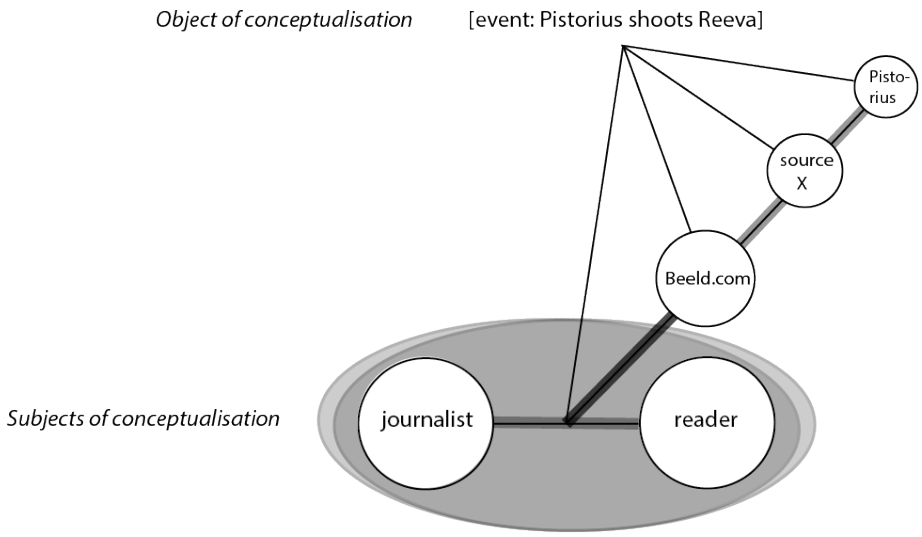


Figure 12 – Sentence (4)

All in all, using the conceptual framework suggested here, and depicted schematically in Figure 4-12, I argue that linguistic elements across different categories, levels of analysis, and languages (lexical units, grammatical and narratological patterns, English, Dalabon) operate along three dimensions: speaker and addressee negotiate ((*x*)-axis) how the common ground should be updated with respect to a particular object of conceptualisation ((*y*)-axis), potentially by inviting the other to view this object of conceptualisation (in part) from the perspective of third parties ((*z*)-axis). In the case of some interaction events this process of updating the common ground involves no third-party perspectives at all (to those interaction events only the first two dimensions are relevant), whereas in other cases a complex thoughtscape is conjured up in the course of this process. Sometimes, the perspectives in these thoughtscapes are embedded into one another (cf. Figure 11 and 12) and sometimes they are related in different ways. For example, third parties can themselves be represented as being interlocutors in a different interaction event (cf. Figure 8), or their viewpoints can form meaningful conjunctions or exhibit causal relations from the perspective of the speaker and addressee. These latter two options have not been explored using examples in this chapter (however, Chapter 6 will feature

several examples). In conclusion, it is worth noting that the suggested conceptual space can accommodate such thoughtscapes comprising conjunct, causally related, or otherwise linked third-party perspectives:

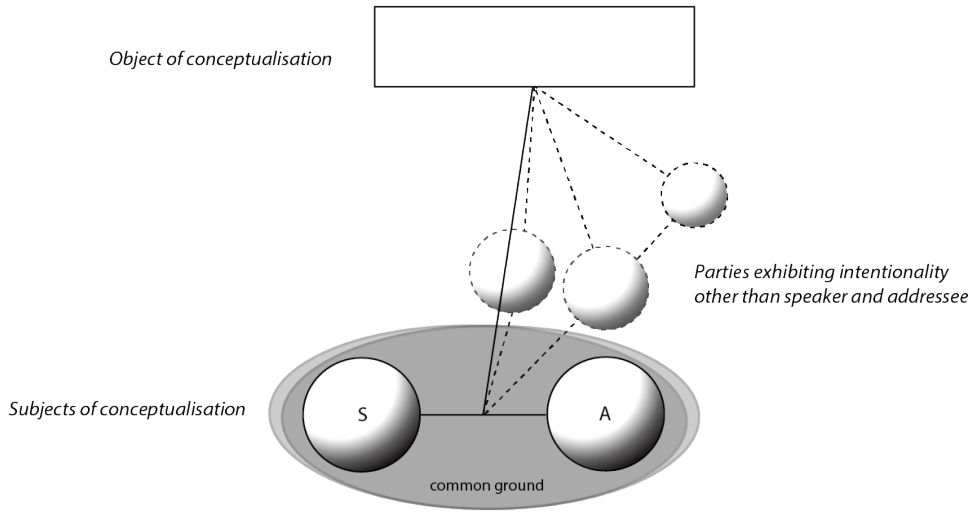


Figure 13 – In the course of some interaction events, a thoughtscape is conjured up taking the form of a network of perspectives related in various kinds of ways. This has been discussion in more detail in Chapter 3 and will be further discussed in Chapter 6. The dashed lines and circles suggest a conjunction between one single perspective and two perspectives exhibiting a form of embedding (e.g. "John believes that X while Mary thinks that Peter doesn't want that X").

5.3.4 Updating the common ground

In this final subsection I will introduce the view of viewpoint coordination as a matter of highlighting, negotiating, and anticipating how individual perspectives *deviate* from the common ground (see again also Clark, 1996). This view will be important throughout the next chapter and be built on in the Conclusion. As one last example, consider the following excerpt of a recorded conversation from the Corpus of Spoken Dutch (CGN), followed by my English translation:

(7) A: oh dan is vandaag Allerheiligen.
B: 't is vandaag Allerheiligen ja. [...]
A: oh oh dan heeft Ella zich denk ik vergist.
want ze dacht dat morgen Allerheiligen was en dan waren de winkels beperkt open
B: ja.
A: nee dat is dat is uh dat is vandaag.

A: oh in that case today is All Saints.
B: today is All Saints indeed. [...]
A: oh oh then I think Ella was mistaken.
because she thought that tomorrow was All Saints and then the shops were only open for a limited period of time
B: yes.
A: no that is that is uh that is today.

Interlocutor A finds out that Ella, a third-party subject not present in the current interaction event, falsely believed that All Saint's Day was tomorrow. The underlying assumption is that knowledge of when this is, is part of the common ground within the cultural-linguistic community of which A, B, and Ella are apparently members. Interlocutor A first opposes Ella's false belief-state to the common ground using the viewpoint package "mistaken" (i.e. *holistically* in terms of Chapter 4), and then further elaborates using the complementation construction "she thought that tomorrow was All Saints" (*compositionally* in terms of Chapter 4). Just as in the example given above of my office mate and me having a misunderstanding (note the viewpoint package!) over whether pointing was directed at the coffee machine or the window blinds, the working out of different perspectives enters the stage in order to figure out a *deviation* from the common ground.

In fact, this is not different with the Pistorius case: the entire thoughtscape hinges on the fact that there are two competing versions of the story (i.e. a crucial discrepancy in common ground) distributed over various third-party subjects. What the interaction as depicted in Figure 11 boils down to is a

journalist negotiating the exact nature of this difference in front of a reader. In Figure 8, depicting the situation of Simon saying to Arran "John assured Mary he would be on time", we see how Simon singles out John's perspective, which is contrasted to both interlocutors' (and Mary's) expectation that he will be late. Recall also once again the case of Shakespeare's *Othello* discussed in detail in Chapter 2: the plot of this play combines multiple scenarios (revenge, a scheming plan, suspected adultery) that imply crucial knowledge differences between the involved parties, inducing a long sequence of negotiations about how various character mindstates deviate from a common ground. I will say a few more words about this point in the Conclusion, after it has been applied in the context of testing intentional-reasoning competence experimentally in Chapter 6.

5.4 Discussion and concluding remarks

Verhagen (2005: 4) argues that "mental coordination" is an essential part of linguistic interaction, and therefore it is to be expected that languages have developed, over the course of their history, special conventionalised signals to support this function, in line with Du Bois' (1985) claim that grammars code best what speakers do most. Verhagen (2005) focuses specifically on words and constructions (besides gestures, facial expressions, and other meaningful elements) which support mental coordination between speaker and addressee, but here I have cast the net wider and also included the marking and coordination of the mindstates of third parties, who may or may not be present at the time of speaking, or who may even exist only in the imagined worlds of thought and fiction. What I have argued is that linguistic items capable of viewpoint coordination serve to highlight and negotiate how individual perspectives *deviate* from the common ground. This reflects an important characteristic of human interaction: instead of starting from individual intentional systems that seek to become "paired", the default is that interlocutors take part in a system of shared intentionality or common ground and negotiate how individual perspectives relate to this.

An important remaining question is how all of this affects our evolutionary story. As pointed out in Chapter 1, it is generally assumed that our ancestors had to reach a certain threshold of intentional reasoning capacity before communication "as we know it" could begin. Indeed, according to Sperber (2000) and Scott-Phillips (2015), the capacity to reason at five orders of intentionality had to predate "proper" ostensive-inferential communication. After all, individuals had to mutually recognise communicative and informative intentions, understanding that the other intends one to see that the other intends one to understand that something is the case. However, the view advocated in this chapter allows for an alternative: I suggest that our ancestors in some way first started to establish forms of common ground, and then developed increasingly sophisticated ways of singling out individual perspectives and ways in which they differed.[85] This process is presently reflected in all kinds of linguistic items being capable of highlighting and negotiating how the perspectives of signallers, addressees, and third-party subjects relate to and, indeed, deviate from the common ground.

Finally, note that this, on an abstract level, is a similar kind of theoretical "move" as the one made by Shultz et al. (2011) regarding early primate social life. Their evidence seems to support a scenario in which individuals *first* started living (c.q. foraging) in groups, and *then* developed increasingly profound dyadic bonds and relationships. In the Conclusion I will integrate this point in Dunbar's framework as set our in Chapter 1. However, before getting there I will apply the developed views to the practice of assessing multiple-order intentionality experimentally in the next chapter.

[85] Note that this is much in the fashion of what Moll and Tomasello (2007) term the "Vygotskian intelligence hypothesis" (cf. Vygotsky, 1978).

Chapter 6

The Mentalising Test Revisited

In several parts of this thesis I have referred to "mentalising tests", versions of the instrument used by experimental psychologists for assessing people's ability to reason about intentional states at various levels of complexity. In the Introduction and in Chapter 1, various key findings of research using these tests were listed, such as the correlations between individuals' mentalising scores and the sizes of their social networks, or the amount of grey matter in particular brain areas. In Chapter 2 a story used in a version of the test was cited (about Emma trying to get an increase in wages with her job at the greengrocer) and an example of a question was discussed (Jenny *wanted* Emma *to believe* that her boss *thought*…etcetera), after which the distinction was made between *situations* involving multiple orders of intentionality and their linguistic *representation* in the form of a proposition or narrative. Chapters 2, 3, and 4 discussed representations of multiple-order intentionality across plays, novels, journalistic discourse, and spoken language. In Chapter 5, I have argued that, when interacting, reasoning about orders of intentionality enters the stage especially when we need to determine how individual perspectives *differ* from shared knowledge or common ground. The present chapter first discusses key issues of mentalising tests in general, in the light of insights from the previous chapters, followed by a detailed investigation of a selection of stories and questions from three mentalising studies.

6.1 Five central conclusions of the mentalising-test paradigm

As discussed in the Introduction and in Chapter 2, in mentalising tests "complexity" has always been conceptualised as the number of embedded intentional states featured in each question, following Dennett's scale of the orders of intentionality (see Chapter 1, Section 1.2.1). As an illustration, consider the following two true/false-statements[86] from the version of the mentalising test used by Brown (more details of her study follow in 6.2 below):

(1) Sam wanted to buy a stamp
(2) Henry knew Sam believed he knew where the Post Office was

After a short story was read out twice, participants had to answer "true" or "false" for twenty of such statements. Ten were "intentionality questions" and concerned intentional states of characters featured in the story; the other ten were "memory questions" and concerned factual details, for example:

(3) The Post Office was closed because it had moved to Bold St

In the case of intentionality questions, the level of complexity for each question is determined by counting the number of embedded intentional states, whereas in the case of the memory questions, complexity corresponds with the number of factual details included in the statement. In this way, statement (1) has complexity level 2, since the participant has to work at second-order intentionality following Dennett's scale: the intentional system (the participant) has to $know_1$ whether or not Sam $wanted_2$ to buy a stamp. Statement (2) has complexity level 4, since the participant has to $know_1$ whether or not Henry $knew_2$ Sam $believed_3$ he $knew_4$ where the Post Office was. In a similar way, statement (3) is a memory question at complexity level 2, since it has two factual

[86] Some of the studies used true/false statements and others (such as the original study by Kinderman et al. (1998) or the recent one by O'Grady et al. (2015) discussed below) presented two alternative statements from which participants had to choose the right option.

elements which have to be checked against the story: *the Post Office was closed*$_1$ because *it had moved to Bold St*$_2$.

The original finding by Kinderman, Dunbar, and Bentall (1998), who used a test comprising five such stories and sets of questions, is that error rates go up steeply at complexity level 6 (when counting as explained above)[87] in intentionality questions, but not in memory questions. The following graph presents this result:

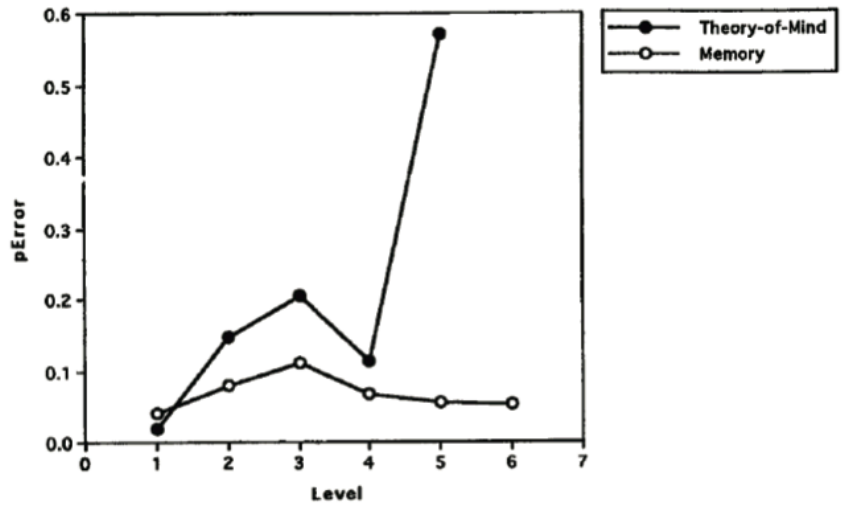


Figure 1 – Graph from the original paper by Kinderman et al. (1998). The proportion of incorrect answers (error rate) is indicated on the y-axis; the level of complexity on the x-axis. Mentalising questions are labeled "Theory-of-Mind"; questions about factual details "Memory"; n=77 participants.

The *asymmetry* between error rates at higher levels in intentionality versus memory questions, along with indications of significant variance in performance *between* participants, as found in this study, were interpreted as

[87] It is important to note that in this paper the number of orders was counted starting from level 0, which means that their level 1 would be referred to as level 2 in terms of the counting method used in this thesis and in by far the majority of cases throughout the literature. In the counting of Kinderman et al. (1998), error rates thus went up at level 5 instead of level 6. The rationale behind the common way of counting is given in Chapter 1, Section 1.2.1: in for example (1) above, the participant has to work at second-order intentionality, since s/he has an intentional state about Sam having an intentional state.

the first evidence for three of the total of five central conclusions of the mentalising-test paradigm that I will distinguish:[88]

conclusion (i) In mentalising questions, mean error rates remain constant or increase to a limited degree with each level of complexity added, until a steep increase takes place at level 6 (i.e. level 5 counting according to Kinderman et al.), suggesting the presence of a natural "limit" at that level.

conclusion (ii) There is significant between-subject variance in error rates at the different levels of complexity, suggesting that some individuals have their "limit" at level 4, some at level 5, some at level 6, and some at an even higher level of mentalising complexity.

conclusion (iii) Performance on mentalising and memory questions is related (participants making more mistakes in memory questions, also tend to make more in intentionality questions), but the results cannot be explained in terms of memory alone: beyond remembering multiple, mutually related details from the story (which is necessary for answering both types of questions), reasoning about embedded intentional states adds an extra challenge, as reflected in the differences in error rates at especially the higher levels of complexity.

Another issue investigated in the paper by Kinderman et al. concerns correlations between scores on the mentalising questionnaire and scores on another test, known as the "Internal, Personal and Situational Attributions

[88] I will refer to research using versions of the discussed method as studies within the "mentalising-test paradigm". The original title of the questionnaire was the "Imposed Memory Task" or "IMT", and it is sometimes referred to using this title across the literature.

Questionnaire" or "IPSAQ". This latter test was designed to measure participants' tendency to attribute negative and positive experiences in social settings to either themselves, others, or situational circumstances. A (strongly) increased tendency to see others as responsible for negative social events is associated with psychopathological disorders (most notably, paranoia). In a healthy subject population, there is nonetheless individual variation in how such attributions are made. Kinderman et al. showed that this variation in part correlates negatively with mentalising performance: those individuals who exhibited less ability to reason about intentional states had a higher tendency to attribute negative social events to others (or, put bluntly, "were somewhat more paranoid"). This was interpreted as the first evidence for what can be identified as the fourth central conclusion of the mentalising-test paradigm:

conclusion (iv) Mentalising performance is relevant to people's actual social lives: an individual's mentalising score tends to be reflected in indicators of this person's social life and general aptitude in the social domain.

The patterns found by Kinderman et al. have been replicated in an array of subsequent studies among different participant populations. These studies used improved versions of the mentalising questionnaire and introduced new measures against which the mentalising scores were tested. In this way, they have yielded additional evidence for the conclusions (i) – (iv), and added a fifth one situated in the neuroscientific realm (see below). The most important additions include findings pertaining to several domains of social and cognitive functioning, further supporting conclusions (ii) and (iv). First, performance at higher levels of mentalising complexity was found to be lost in patients suffering from bipolar disorders (Kerr, Dunbar, and Bentall, 2003). Secondly, estimates of people's personal social network size turned out to be associated with their mentalising scores. Stiller and Dunbar (2007) found that people who exhibited higher mentalising performance, on average tended to have a larger "support clique" (defined as the innermost circle of social contacts, from which

one receives emotional support).[89] Thirdly, studies using a version of the mentalising test adapted for children indicated that higher scores were associated with higher social competence as independently rated by their teachers (Liddle and Nettle, 2006). Moreover, associations were found between mentalising performance and personality traits (Nettle and Liddle, 2008; Van Duijn et al., 2014), schoolgrades (Van Duijn et al., 2014), empathy, and aspects of executive functioning (Launay et al., 2015).

The fifth and final central conclusion of the mentalising-test paradigm can be formulated as follows:

conclusion (v) Mentalising performance is related to brain size, in particular to the amount of grey matter in the orbital prefrontal cortex in humans.

This conclusion forms an important background to the idea, discussed earlier at the end of Chapter 1, that our hominid ancestors gradually, over many generations, evolved an increasing capacity for mentalising, going from a limit at around complexity level 2 in our last common ancestor with chimpanzees and bonobos, through a limit at level 3 in homo erectus, one at level 4 in archaic humans, and eventually a limit at level 5 or 6 in anatomically modern humans (see Chapter 1, Section 1.3.6 above). This idea rests on the pillars of two findings. Firstly, monkeys with smaller brains seem capable of only first-order intentional attributions, whereas monkeys and apes with relatively bigger brains can (under certain circumstances) handle second-order intentionality (see e.g. Call and Tomasello, 2008). Secondly, as expressed in (v), there is evidence in humans that the subjects in test populations who perform better at mentalising tests, have more brain volume in the orbital prefrontal cortex, an area associated with various aspects of social functioning (Powell et al., 2010; Lewis et al., 2011). More precisely, the claim is that the amount of grey matter in the orbital prefrontal cortex is positively correlated with social network size,

[89] Even more interesting than the correlation with mean support clique size might actually be the possibility that mentalising competence imposes a *limit* on an individual's maximum possible support clique size—Stiller and Dunbar find some support for this suggestion in their data (see 2007: 98-100). The correlations of mentalising scores with estimates of mean social network size were replicated by Lewis et al. (2011), Powell et al. (2012), and Powell et al. (2014).

and that this relation is *mediated* by mentalising competence (Powell et al., 2012; Powell et al., 2014). Note that this comes down to a within-species version of the social brain hypothesis: bigger and more powerful brains allow for the management of larger and more complex social networks not only between different species of primates, but also between human individuals.[90]

All in all, the current status quaestionis is that individuals having relatively more brain volume in particular areas relevant to social functioning (through genetic or developmental factors, or both), possess some social competences that can be measured using the mentalising test. As a consequence of these social competences they can maintain a larger and more complex social network. What those social competences measurable with intentionality tests entail *precisely* is unclear, but they correlate with particular psychometric measures (causal attribution of negative and positive social events) and indicators of traits such as short-term memory performance, personality, empathy, and, potentially, executive functioning. I will get back to this interpretation below and in the Conclusion.

6.2 Testing mentalising competence

For non-human species, a broad range of behavioural tests have been developed to investigate their abilities and limitations in reasoning about intentional states. Examples include (to name just a few) interpreting informative cues (e.g. Premack and Woodruff, 1978; Tomasello, 2008), hiding and tracking food in competition with others (e.g Hare et al., 2006, for chimpanzees; Clayton and Emery, 2004, for corvids), or working together to gain access to food (De Waal, 2005; Yamamoto, Humle, and Tanaka, 2012). Such

[90] The same trend has been found in macaques, although evidence there seemed to indicate that the increase in grey matter in relevant areas was the *result* of living in a larger social network rather than a precondition (see Sallet et al., 2011). This interpretation seems also feasible in humans: living in more complex social environments (such as a larger personal social network) could increase social capability, which is then reflected in the amount of grey matter one has in the orbital prefrontal cortex and one's scores on a mentalising test. Note that these scenarios (more complex social life>more grey matter; more grey matter>more complex social life) are not mutually exclusive: "boot-strapping" or co-evolution is a likely possibility.

tests have broadly been combined with observations in the wild and other forms of "anecdotic" evidence (in the non-pejorative sense; see e.g. De Waal, 2005; Boesch, 2005).

For normally developed human adults, as well as for infants and several "special" subject populations (such as individuals suffering from psychopathological disorders), an even greater variety of tests attempting to provide insight into intentional reasoning abilities have been used (see Apperly 2011 for an overview). Roughly, these tests can be classified as follows:

	explicit/reflective	implicit/behavioural
1st- and 2nd-order	A	B
3rd-order and above	C	D

Table 1

Categories A and C include tests that are mediated by descriptive language of a kind one would not normally find in interaction. Such language usage is very similar to what is in Chapters 1 and 2 referred to as "propositional" or "formal". For example, the proposition "Mary intends that John believes that it is raining outside" is quite different in nature from the actual (linguistic and/or non-linguistic) behaviour one can expect Mary to exhibit in the described situation, which would probably consist of saying "Hey, it's raining", "Be prepared to get wet", or just handing John an umbrella (see also Section 6.2.3 below on this point). In a similar way, classic false-belief tasks (such as the original version of the Sally-Ann test, see Apperly, 2011: chapter 2) ask of participants to *reflect* on a situation using explicit descriptions of mindstates and behaviour, rather than participating in such a situation oneself, which makes them fall under A in the above matrix. For the same reason, classic mentalising tests (following Kinderman et al., 1998) fall under C, since participants have to read a story describing a social situation and then are asked to reflect on this situation by judging statements as true or false.

By contrast, implicit/behavioural tests, indicated by B and D in the above matrix, ask of participants to take part in a (controlled) social situation. Examples of such tests include those in which children have to provide a

particular kind of help to an experimenter who tries to solve a task (after which conclusions can be drawn about their ability to judge the experimenter's goals; see Tomasello, 2008) or a "Schelling game" in which participants have to ask themselves what the other would do, what the other will think they would do, and so on. Note that in practice, A/C and B/D are extremes on a scale rather than clear-cut categories. For example, Baron-Cohen's (2001) Reading the Mind in the Eyes-test could be located somewhere in the middle, as it does use descriptive language to label the moods that participants are supposed to pick up from the pictures they are shown (more A), but also involves the "implicit" act of reading the eyes (more B).[91]

In what follows, the focus will be on three studies that can be located on the imaginary scale ranging from C to D. One is a "classic" mentalising test, close to the Kinderman et al. (1998) original. The other two broadly maintain the format, but replace stories and/or questions by dialogues or movie clips, marking a move from C (somewhat) towards the direction of D in the matrix above. I will refer to each study using the surname of its primary investigator. The studies can be outlined as follows:

Brown — This is a classic mentalising study (C in the matrix above) that was performed using a questionnaire adapted from the original by Kinderman et al. (1998). It was run by Rachel Brown as a pilot for subsequent neuroimaging studies (leading to the publication of Lewis et al., 2011) and remained unpublished. It featured 25 participants (18 female; age 21-76) to whom the stories were read out twice, after which the same was done with each question (they could not see the story or the questions in written form so they had to go by their ears). Each story was followed by around 20 questions, typically 10 intentionality and 10

[91] Note that this scale is also to some degree meaningful in the case of tests for non-human animals. Even though such tests are of course non-linguistic, they can still be more reflective or more behavioural: for example, Premack and Woodruff's (1978) chimpanzee Sarah had to choose which tool use could solve a task using video recordings (which is, relatively, more A), whereas chimpanzee's in an experiment by Yamamoto et al. (2012) had to actually hand the right tool to a peer themselves (which is B).

memory ones in random order. Participants had to answer "true" of "false" using two buttons after each question.

Haddad

This study, run by Anneke Haddad, had the same procedure as Brown's, but stories were replaced by dialogues that were recorded using different voices for each character (thus arguably moving slightly on the scale from C to D, given that the dialogues mimic real-life interaction more closely than descriptions of social interaction do). However, the questions were still presented in the classic format. There were two participant groups: adolescents and adults. Slightly different versions of the dialogues were used to attune to each participant group: "colleague" was replaced by "class mate" for adolescents, and so on. The results are currently under review; I thank Anneke Haddad for permitting me to use her study here.

O'Grady

This study, run by Cathleen O'Grady, was recently published in *Evolution and Human Behaviour* (O'Grady, Kliesch, Smith, and Scott-Phillips, 2015) and comprised four conditions: (i) stories acted out and presented using movie clips, followed by a series of pairs of alternative continuations of the story also in acted movie clips, one of which was "true" and the other "false" (called the "implicit-implicit" condition by the authors); (ii) acted movie clips followed by pairs of alternative propositions read out by one "actor" facing the camera (implicit-explicit); (iii) stories read out by one actor facing the camera, followed by pairs of alternative continuations of the story in acted movie clips (explicit-implicit); and (iv) stories read out followed by propositions read out (explicit-explicit). Participants were allowed to watch each item as often as they wanted, but could not go back once

> they had gone to the next item. The conditions can be placed on the scale from C to D in reverse order: (iv) is closest to C, (iii) and (ii) are in between, and (i) is relatively closest to D. However, none of the conditions is "D proper", given that even the implicit-implicit condition, featuring movie clips mimicking real-life interaction to some degree, yields a reflective rather than a behavioural test. Participants were 66 students (41 female) who declared that they did not know the actors in any of the movie clips.

Below I will discuss selected examples from these studies; more questions and stories can be found in the Online Appendix (see note 95). Given that Haddad's and O'Grady's studies are derived from the type of study represented here by Brown, I will discuss this latter one in detail and use it as a benchmark against which the other two can be compared.

6.2.1 Narratives and propositions

Brown's study is a good representative of what could be called the "classic" mentalising study. It used a questionnaire close to the Kinderman et al. (1998) original, which was also used (abstracting from some various smaller revisions) in many other studies over the years, for example: Stiller and Dunbar (2007), Powell et al. (2010), Lyons, Caldwell, and Shultz (2010), Lewis et al. (2011), Powell et al. (2012), Powell et al. (2014), and Launay et al. (2015). I will first discuss some issues I consider to be general for al these studies, using the first story and question set of Brown's questionnaire.

(4) WHERE'S THE POST OFFICE?

> Sam wanted to find a Post Office so he could buy a Tax Disc for his car. He was already late buying one, as his Tax Disc had run out the week before. Because traffic wardens regularly patrolled the street where he lived, he

was worried about being caught with his car untaxed. As Sam was new to the area, he asked his colleague Henry if he could tell him where to get one. Henry told him that he thought there was a Post Office in Elm Street. When Sam got to Elm Street, he found it was closed. A notice on the door said that the Post Office had moved to new premises in Bold Street. So Sam went to Bold Street. But by the time he got there, the Post Office had already closed. Sam wondered if Henry, who was the office prankster, had deliberately sent him on a wild goose chase. When he got back to the office, he asked another colleague, Pete, whether he thought it likely that Henry had deliberately misled him. Pete thought that, since Sam had been anxious about the Tax Disc, it was unlikely that Henry would have deliberately tried to get him into trouble.

After hearing the story twice, participants were presented with the questions listed below (answer and level are added here; during the test questions were presented in mixed order). They received the instruction: "do not guess, if you think you cannot answer the question on the basis of (what you remember from) the story, choose 'false'".

	Intentionality questions	**Answer**	**Level**
1	Sam wanted to buy a stamp	F	2
2	Henry wanted to play a trick	F	2
3	Henry thought that Sam knew he was a prankster	F	3
4	Pete suspected that Henry was playing a prank on Sam	F	3
5	Henry knew Sam believed he knew where the Post Office was	T	4
6	Sam thought Henry knew he wanted a Tax Disk	F	4
7	Sam thought that Henry knew the Post Office was in Bold Street and hence that Henry must have intended to mislead Sam	T	5
8	Sam believed that Pete thought the Post Office was in Elm Street and hence that Pete must not have intended to mislead Sam	T	5
9	Pete wanted Sam to know that Henry believed that the Post Office	T	6

	was on Elm Street and hence did not intend to mislead him		
10	Pete wanted Sam to know that he believed that Henry had intended not to mislead him	T	6
	Memory questions		
11	Sam needed a Tax Disc from the office	F	1
12	The Post Office was closed and Sam's insurance had run out	F	2
13	The Post Office was closed because it had moved to Bold St	T	2
14	The Post Office in Elm St. had a notice on the door	T	2
15	Sam left Bold Street, then went to the office and spoke to Pete	T	3
16	Sam found the Post Office closed and couldn't buy a tax disk for Pete	F	3
17	Pete, the man who worked at the same place as Henry, and who knew that Henry was the office prankster, was Sam's cousin	F	4
18	Sam asked Henry, and did not ask Pete or the traffic wardens, about where the Post Office was in order to buy a Tax Disk	T	4
19	Sam who worked with Pete and Henry did not know where to buy a Tax Disk because he was new to the area	T	4
20	Henry, the man that Sam spoke to about where to buy a Tax Disk after he realized he needed to buy one soon, was a colleague of Pete's	T	4

Table 2

Looking at this test, several observations can be made. First of all, some of the formulations of the questions are quite puzzling: it could well be that the participants did understand the story correctly, even remembered the relevant details about who-knew-what, who-wanted-what (etc.), but then got lost when dealing with the questions. As discussed in Chapter 2, propositions of the form used for the higher-order questions (such as questions 7, 8, 9, and 10) do not exist in the "wild", so language users cannot rely on experience when assessing

them.[92] The stories communicate the same information in a somewhat more natural way: in fact, they offer another demonstration that in natural communication "narrative takes over" when more than three perspectives are involved. This leads to the first general issue:

issue (i) Classic mentalising tests use *narratives* to present a social situation, but use *propositions* to present the questions. Especially above complexity level 3, such propositions are a very unnatural way of representing intentional states in discourse. With the analysis from Chapter 2 in mind, I suggest that the unnatural presentation of the questions can be a factor limiting performance on especially higher levels of complexity.

On top of this, the propositions used in the questions vary in complexity, measured in terms of the number of *embedded* mindstates they present. However, in the stories mindstates may not only be embedded into one another, but also connected in different ways. As an example, consider the following two sentences from the story cited (4) above:

(5) Sam wondered if Henry, who was the office prankster, had deliberately sent him on a wild goose chase. When he got back to the office, he asked another colleague, Pete, whether he thought it likely that Henry had deliberately misled him.

Looking at this passage in the way used to analyse the narrative texts in Chapter 3 and the news-paper excerpts in Chapter 4, one must conclude that it

[92] An indication that such sentences are very infrequent or even non-existent at least in spoken discourse is that the Corpus of Spoken Dutch (CGN; 10 million words) features not a single sentence with four or more embedded intentional clauses (see also Chapter 2). Note that from this indication that these sentences are idiosyncratic in the context of everyday interaction it does not automatically follow that the test cannot be used to assess some aspects relevant to such interaction—quite generally, tests do of course not always have to mimic their target subject as closely as possible to be valid. However, when interpreting test results it is crucial to know in what respects the test differs from "real life", and in what follows I suggest that, in the case of studies investigating mentalising using the test described in this section, this has not always been in clear focus.

coordinates a thoughtscape of mutually interlinked, but not necessarily embedded, intentional states. First of all, there is a narrator providing insight in the perspective of Sam. Using a form of indirect thought ("wondered if…"), the reader (or, in the experimental setup, hearer) is informed that Sam wants to know whether or not Henry had deliberately sent him to the wrong location, in order to play a prank. Within the scope of Sam's thoughts, there are thus two alternative mindreads of Henry's intentions: either Henry intended to provide the right location but did not know that the post office had moved, or he did know that the post office had moved, but intended to provide the wrong location because he thought that this would be funny. Eventually, Sam checks with Pete what he thinks Henry had intended. Readers end up with the knowledge that Sam still has two alternative mindreads of Henry to choose from, plus Pete's opinion on which one is the most likely.

Put more compactly, the second issue is:

> issue (ii) Narrative language usage features all kinds of cues that prompt and mutually coordinate intentional states of characters. As the analysis in Chapter 3 showed, a thoughtscape emerging in this way is easily misrepresented by propositions featuring only embedded mindstates. This suggests a *structural discrepancy* between the nature of the complexity presented in the stories and in the questions.

Related to this, an observation that can be made repeatedly in especially higher-order questions in Brown's study is that the chains of embeddings are "broken". Consider question 7 as an example:

> 7. Sam thought that Henry knew the Post Office was in Bold Street **and hence** that Henry must have intended to mislead Sam[93]

[93] Given that the questions from Brown's study are already numbered in Table 2, when citing them I will not continue my regular numbering between brackets: (1), (2), and so on.

The method of schematic depiction of "narrative spaces" introduced in Chapter 3 is once again a useful tool here. The following configuration of narrative spaces can be drawn on the basis of question 7:

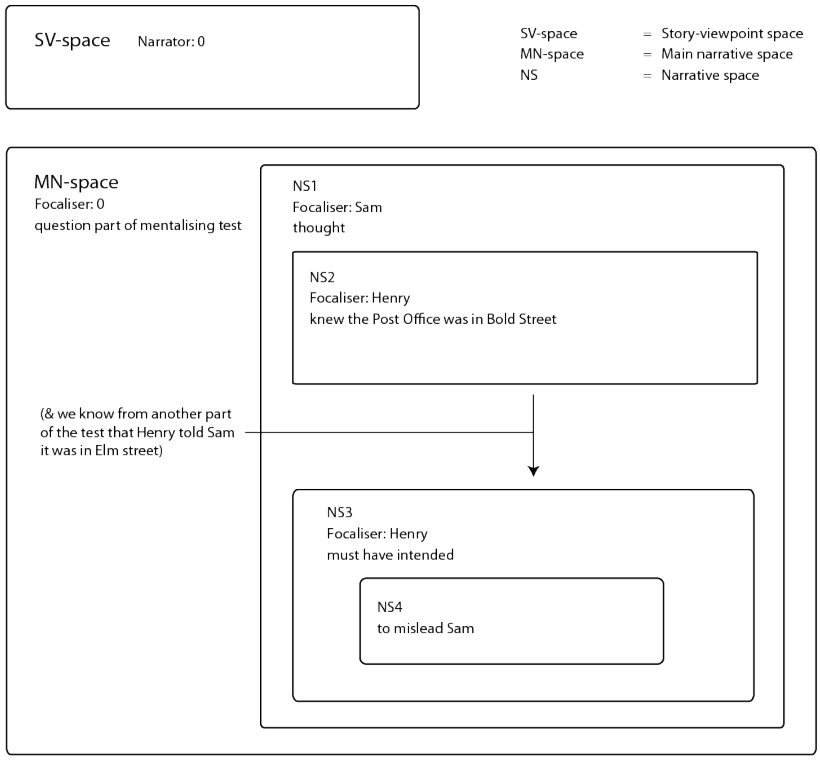


Figure 6 – Schematic depiction of the narrative-spaces configuration prompted by question 7. NS1 and NS2, together with the background knowledge gained earlier when reading the story, work as premises for the conclusion drawn in NS3 and NS4. Sam is the main focaliser in NS1 and all spaces within this space (NS2, NS3, and NS4); Henry is an embedded focaliser in NS2 and NS4. Note that NS4 contains a viewpoint package ("mislead") that could be unpacked into further spaces (see Dancygier, 2012, and Chapter 3 above for more details about narrative spaces).

Question 7 is not composed of a continuous string of four embeddings, but instead of a proposition exhibiting two embeddings ("Sam thought Henry knew the Post Office was in Bold Street"), coupled to another proposition ("Henry must have intended to mislead Sam") using a combination of connectives

marking causality ("and hence").[94] This is different in question 10, in which the chain is unbroken:

> 10. Pete wanted Sam to know that he [Pete, MvD] believed that Henry had intended not to mislead him [Sam, MvD]

Here the following schematic depiction can be drawn:

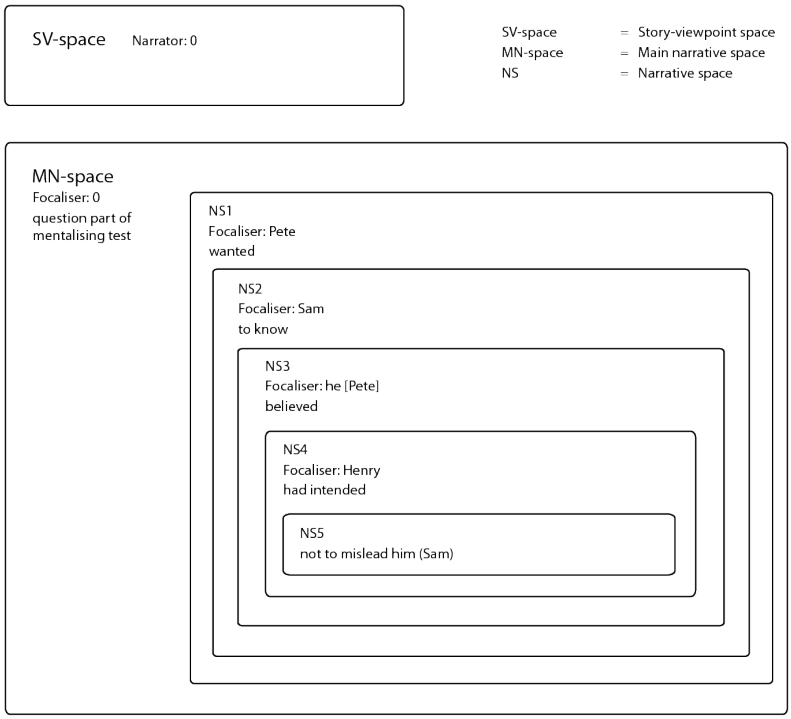


Figure 7 – Schematic depiction of the narrative-spaces configuration prompted by question 10. As in the case of Zunshine's paraphrase of *Mrs Dalloway* discussed in Chapter 3, each narrative space is embedded into the former one.

[94] Note that alternative interpretations of how the question should be read can lead to slightly different narrative-spaces configurations. The one drawn here would in fact be expressed more naturally by the sentence: "Sam thought that Henry knew the Post Office was in Bold Street and hence that Henry must have intended to mislead *him*"). Usage of "Sam" instead of "*him*" seems to suggest that the second proposition (NS3 and NS4) is not what Sam thought (so not part of NS1), but instead added by an external observer/narrator (hence best drawn as part of the MNS in Figure 2). However, this interpretation is again countered by the word "must" (suggesting an inference on Sam's part), meaning that the interpretation drawn here seems to fit best after all.

A similar broken-chain structure as in question 7 can be found in the questions 8 and 9 cited in Table 2 above, as well as in a handful other questions throughout the questionnaire.[95]

The paper presenting the results of O'Grady's study (see O'Grady et al., 2015) begins with a critical discussion of the questionnaire used in most previous studies, of which Brown's study discussed here also uses a version. The authors object to the use of broken-chain questions, stating that instead of testing a "single metarepresentational unit" of higher-order complexity, these test the conjunction of multiple lower-order mindreading tasks that can be processed as separate chunks. This means that participants can possibly use a short-cut to determine the right answer: in some cases they can check the separate parts against the "reality" of the story without ever having to considering the statement as a whole (more about this below). O'Grady et al. argue that including broken-chain questions therefore boils down to a methodological flaw. I agree with their critical stance in as far as *comparing* questions with and without broken chains is concerned: it may be methodologically tricky to assume that a, say, fifth-order question with a broken chain exhibits the same complexity as one without a broken chain. However, regarding the general *validity* of both types of questions I come to a different conclusion. Whereas O'Grady et al. suggest to avoid broken-chain questions because they do not test "real" higher-order mindreading, my analysis suggests they may in fact be better at testing how well participants understood the relationships between intentional states contained in the story. As pointed out in issue (ii) above, the story presents the intentional states not as being just embedded, but instead prompts a thoughtscape of intentional states that are interconnected in all kinds of ways. From that perspective, a question such as 7 above exhibits a structure that is more "realistic", compared to the narrative presentation of events in the story, than does question 10. Broken-chain questions, as it were, "burst" out of the artificial straitjacket of embedding-only propositions above a certain level of complexity and adopt a structure that leaves more space for expressing the viewpoint complexity

[95] The broken-chain/narrative structure can be found in questions 29, 30, 31, 48, 50, 67, 68, 69, 70, and 85. See the Online Appendix for the full questionnaire at http://liacs.leidenuniv.nl/~duijnmjvan/TLM/Appendix, password "thelazymindreader".

contained in the story in a nuanced way—once again narrative seems to be taking over.

6.2.2 Packages and inferences drawn from the common ground

Some of the viewpoint layers are fully spelled out in the story, i.e. prompted *compositionally*, to use the term introduced in Chapter 4. For example, "Sam wondered if Henry [...] had deliberately sent him on a wild goose chase" features the compositional construction of Sam's viewpoint using the verb "wondered" and the complement "if Henry [...] had deliberately sent him on a wild goose chase". However, there are additional layers contained in the story that are coordinated *holistically*: following the conditions defined in Chapter 4, the word "mislead" is a viewpoint package adding extra viewpoint layers from which the related content can be seen. The same goes for "prank(ster)" and arguably for "deliberately" and "sending someone on a wild goose chase". The entire possible mindread of *Henry knowing where the post office is, but deliberately telling Sam the wrong location because he thinks this is funny*, is nowhere spelled out but added holistically by the combination of these cues. It is possible for the reader or hearer to unpack (or *decompress*; cf. Chapter 4, Section 4.3.6) this mindread into single constituent propositions (as I just did in italics), but this is not necessary for following the story as such: as argued in Chapter 4, readers or hearers can take a viewpoint package on board holistically, and integrate only its relevant implications in their understanding of the situation.

The third and fourth issue I want to point out pertain to these dynamics of unpacking (or not unpacking) the situations and events presented by the story into single viewpoint layers. Consider again the passage cited in (5) above, this time in relation to question 2:

> 2. Henry wanted to play a trick

There is no direct evidence that Henry wanted to play a trick, but also no conclusive evidence to the contrary. Or more precisely: it is not written explicitly in the story that Henry wanted to play a trick (i.e. there is no compositional construction of this viewpoint), nor are there any other (holistic)

cues provided from which the inference can be drawn that Henry wanted to play a trick. All we know is that *Pete thought* that Henry did not want to play a trick, given Sam's anxiety about the tax disc. Although Pete's view sounds reasonable (and leads to the right answer according to the test: "false"), strictly speaking participants cannot know the answer to this question. They are best advised following the instruction (cited before Table 2 above) to choose "false" whenever the text provides no conclusive evidence. Out of Brown's 25 participants 23 indeed chose "false", thus answering correctly according to the test.[96]

A similar analysis can be made in the case of question 3:

3. Henry thought that Sam knew he was a prankster

There is no clear answer provided to this question either. In the story, there is a narrator inserting the comment "Henry, who was the office prankster" when reporting Sam's thoughts (see excerpt (5) above), suggesting that this is shared knowledge or common ground for everyone working at their office. In this sense, in the "reality" of the story, it is quite likely that Henry *does* know that Sam knows that he (Henry) is a prankster—after all, this is what he can derive from this being common ground. Question 3 thus presents a thought that Henry theoretically *could* have had, but, realistically speaking, only *will* have had if a context occurred in which it was relevant for him to derive this information from the common ground. Such contexts would be, for example: him *understanding* that Sam is anxious about the tax disc, and therefore *anticipating* that Sam might not trust him straight away, given that he is known as the office prankster; or: the context of Sam coming back from his failed attempt to buy a tax disc and confronting Henry himself (instead of Pete) with the suspicion that he had deliberately misled him, after which they could talk about the *misunderstanding* that occurred because Henry *had not anticipated* that Sam knew that he was the office prankster. However, the story features no evidence for any of these contexts.

[96] More information on the error rates can be found in the Online Appendix, see note 95.

All in all, question 3 presents viewpoint layers that Henry *could have unpacked* if there had been a context requiring this. Given that there is no conclusive evidence of the presence of such a context, the answer should be "false" following the same line of reasoning as with question 2. This is indeed what 24 out of Brown's 25 participants chose, thus answering correctly according to the test.

So far so good—now consider question 5:

> 5. Henry knew Sam believed he knew where the Post Office was

From the story it can be concluded that Sam did believe that Henry knew where the Post Office was, but there is no conclusive evidence whether Henry did or did not know that Sam believed this. The answer can again be "true" in theory, given that Henry had a conversation with Sam in which he told him where the Post Office was, and under normal circumstances Henry should be able to draw the inference that Sam believes what he told him. However, again there is no evidence of a context in which Henry would indeed have drawn this inference, so following the same line of reasoning as with questions 2 and 3 the answer should be "false". Yet according to the test the correct answer is "true", which is what 17 out of the 25 participants went for. The "failure" of the other 8 participants to provide the "correct" answer can be due to their inability to handle the fourth-order-intentionality proposition featured in question 5, but also to the fact that they followed the instruction to choose "false" whenever the story provides no conclusive evidence for the existence of particular intentional states, either by constructing them explicitly/compositionally or by providing a context from which they can be inferred.

Two further issues can be formulated with this analysis in mind:

issue (iii) Following the same line of reasoning leads sometimes to a correct and sometimes to an incorrect answer, suggesting that factors other than the amount of orders of intentionality involved in a question (which is of course the target variable of the test) can quite easily interfere with error rates produced by participants.

It is clear that the tests can be improved by checking the questions one by one for such inconsistencies (and some authors have done so, as examples below will show).[97] In addition, it may be advisable to add a third answer option apart from "true" and "false", which participants are supposed to tick when a question cannot be answered on the basis of evidence from the text. This avoids at least some of the ambiguity between cases where participants have failed to process the intentional reasoning task and those where they have failed to apply the instructions correctly.

A more important and fundamental issue, however, has to do with the unpacking of viewpoint layers where this is not relevant in the context of the story. I have above discussed this for the questions 3 and 5, and question 10 contains another example. From the story we know that Pete and Sam had a conversation about what happened to Sam, and that "Pete thought that [...] it was unlikely that Henry would have deliberately tried to get him into trouble". Question 10 asks whether "Pete wanted Sam to know that he believed that Henry had intended not to mislead him" (the answer is indeed "true"). The core issue questioned here is really whether Pete believed that Henry intended to mislead Sam, which is a fourth-order problem. That Pete wanted Sam to know this, follows logically from the fact that they have a conversation. In other words, the first two orders of the question in fact "unpack" what is naturally given in the discourse situation. In Chapter 3, I have argued that Zunshine unnecessarily starts to count from the author, suggesting that Woolf *intends* her readers to *believe* that Richard Dalloway *thinks*, and so on. However, just as these first two layers normally do not have to be processed by readers of a Woolf novel, I suggest that we are normally not concerned with processing that a speaker *intends* the hearer to *understand* that he *thinks*, and so on. In Chapter 5 I have argued, following Clark (1996; 2006a), that such viewpoint layers belong to the (infinitely large) category of inferences that can be drawn from the common ground. When Sam and Pete have a conversation about what Henry

[97] Also, it must be noted that the outcomes of the tests are averages produced by multiple questions (usually between five and nine) at each level of complexity and by mostly quite substantial samples of participants, which means that such inconsistencies are at least to some extent balanced out as part of regular "error variance" for a test like this.

wanted, it can be inferred that Pete wants Sam to know what he (Pete) thinks about what Henry wanted, just as it can be inferred that Pete wants Sam to know that he (Pete) wants Sam to understand what he thinks about what Henry wanted, and so on. Normally, such inferences are not drawn, since they state obvious truths in a complicated way, without adding anything to what both interlocutors consider to be common ground. However, as said above with respect to questions 3 and 5, it is possible to think of exceptional contexts in which drawing such inferences can be useful. Most notably, this is the case whenever it turns out that interlocutors do not understand each other or are, as it were, not "on the same page". For example, imagine the following conversation between Pete and Sam:

(6) Pete: Henry may be a prankster, but above all he is an empathic person.
Sam: What do you mean?
Pete: I want you to understand that I think Henry did not want to deliberately mislead you, given your anxiety about being too late.

Here, Pete first tries to share his thoughts in an indirect way, expecting that Sam will draw his conclusions on the basis of the information that Henry is an empathic person. However, when Sam makes clear that he does not know what to do with that information in this context, Pete "unpacks" and makes explicit what he wanted Sam to understand.

There is no evidence in the story for a situation in which Pete and Sam are not on the same page, which means that it is unlikely that any of them needs to bother about unpacking the discourse situation into separate viewpoint layers. Of course this does not mean that it is impossible for participants to do this when answering the questions. However, when looking at the questions I think it is important to make a distinction between viewpoint layers that are in some way relevant to the characters in the story and the development of the story's plot, and viewpoint layers that are "generated extra" by unpacking layers that would normally be obvious and/or unnecessary.

All in all, the fourth issue can be formulated as follows:

issue (iv) Participants are asked to reflect, *in the same way*, on:
- viewpoint layers that are relevant to the characters and the development of the plot (whether or not these are spelled out compositionally or cued holistically); and
- viewpoint layers that can *in principle* be inferred from the story, but do not have such relevance.

Potentially, including obvious though irrelevant viewpoint layers in the questions leads to confusion about whether the answer should be "true" or neither-true-nor-false, and thus "false" (see analysis of questions 2, 3, and 5 above). In addition, it may also introduce pseudo-complexity: in question 10 a fourth-order problem is preceded by two obvious layers, which in fact come "for free" with the information that Pete and Sam are having a conversation; it is irrelevant for the interlocutors to reflect on this, neither is it relevant for the development of the plot. It is unclear how a question like 10 compares to a question staging six viewpoint layers that do have such relevance.

6.2.3 Judging facts and intentional states

Another point is that judging whether a factual statement is true or false is *conceptually* a different task from judging whether an intentional statement is true or false. Being so-called control questions, the factual questions clearly have to be different—however, the problem may be that they are too different. Ideally, the only difference between intentional and factual questions would be that the first concern intentional states and the latter do not. This is not the case in Brown's study: another important difference here concerns what I have referred to as "(in)transitivity" in Chapter 1, Section 1.1.2. This property affects strategies available for assessing statements for truth-value, either with respect to "reality", or, in this case, with respect to a story: the "intransitive" nature of embedded intentional states make sure that participants have to process the statement as a whole, whereas the "transitive" strings of factual statements can be checked against the reality of the story element by element. As a consequence, there are often "short-cuts" to the answer available in factual memory questions. Consider the following two questions from Brown's study:

> 4. Pete suspected that Henry was playing a prank on Sam
> 16. Sam found the Post Office closed and couldn't buy a tax disk for Pete

The answer to both questions according to the test is "false". As a participant, in order to determine this for question 4, one has to think about what Pete thought about Henry's intentions—a task in which all elements of the statement are somehow involved. However, seeing that 16 is false is a lot easier: all one needs to know is that Sam was not going to buy something for Pete but for himself.

This means that *in theory* there is a structural gap between the complexity of intentionality questions and factual questions. In practice, as pointed out above, a few of the intentionality questions in at least Brown's questionnaire exhibit broken chains of embedding, sometimes also enabling short-cuts for participants (I will say a few more words about this in the next section). The fifth general issue can thus be formulated as follows:

> issue (v) Questions exhibiting unbroken chains of embedding have to be processed as a whole, whereas in questions containing conjunctions and/or causal links each constituting element can be checked against the story separately. Given that the first category contains only (some of the) intentionality questions and the latter all factual memory questions (and the rest of the intentionality ones), this may have affected the observed difference in performance on both types of questions.

A similar argument is put forward by O'Grady et al. (2015) in their critical discussion of the existing mentalising questionnaire. They make the general point that a part of the intentionality and factual questions of the classic mentalising questionnaire (in this chapter represented by Brown's study) can be answered without processing the entire statement. In addition, they raise two specific objections regarding the factual control questions. First, they demonstrate that the intentionality questions use syntactically more

complicated sentences by counting the number of embedded clauses (they report a significant difference between the overall median being 0 versus 2 in factual versus intentionality questions). Second, they argue that the factual questions are inappropriate controls in the first place, since they do not involve *conceptual* embedding. They suggest to eliminate all possibilities for short-cutting, and to match the syntactic and conceptual complexity of intentionality and factual questions at every level by making use of "non-mental recursive concepts", such as relationships of possession or localisations in space and time. They implement these suggestions in their version of the mentalising test. Consider the following three example questions from their questionnaire ((7) is a mentalising question, (8) and (9) are factual memory questions; participants had to choose between alternative options A or B):

(7) A. Victor knows that Amy knows that Sheila intends that John thinks that she hasn't realised that he likes her.
B. Victor doesn't know that Amy knows that Sheila intends that John thinks that she hasn't realised that he likes her.
(complexity level seven[98]; see story 4, intentionality question 6 from the online supplementary material of O'Grady et al., 2015)

(8) A. Michelle and Nick started dating after a walk in the park, when Nick was tipsy, in the afternoon, on November 22nd, before Thanksgiving.
B. Michelle and Nick started dating when Nick was tipsy, during a walk, in the morning, on November 29th, after Thanksgiving.
(complexity level six; see story 2, control question 5 from the online supplementary material of O'Grady et al., 2015)

(9) A. Shaun is Sheila's supervisor Mike's boss's PA John's best friend's girlfriend Amy's brother.

[98] Note that O'Grady et al. (2015) start counting from level zero, as do following Kinderman et al. (1998). According to their study, a participant judging what a character in the story believes is working at first-level intentionality (instead of second-order as counted in most other studies and throughout this thesis), so they would refer to (7) as being level six. See also note 2 above.

> B. Shaun is Sheila's best friend Mike's supervisor's boss John's PA's girlfriend Amy's brother.
> (complexity level seven; see story 4, control question 6 from the online supplementary material of O'Grady et al., 2015)

O'Grady et al. argue that in all three questions both the concepts and the syntax are recursively embedded. This is clearly the case in the intentionality question in (7), but how about the memory questions in (8) and (9)? In a way, it is true that in (8) the walk in the park is *conceptually* embedded in Nick being tipsy, which is again conceptually embedded in an afternoon, embedded in the day November 22nd, embedded in the period before Thanksgiving. In some way, it can also be argued that the syntax of clause(s) indicating when Michelle and Nick started dating exhibits a recursive structure: a noun phrase embedded in a noun phrase, embedded in yet another noun phrase, and so on ([a walk in the park, [when Nick was tipsy, [in the afternoon, [on November 22nd, [before Thanksgiving]]]]]).[99] However, if the aim is to match the form of intentionality questions as closely as possible, this type of recursion does not do the job: as I have argued in Chapter 1, Section 1.1.2, a distinctive feature of embedded intentional states is that they exhibit "intransitivity": "A thinks that B thinks that C thinks that X" means something quite different than does "A thinks that C thinks that X", and if the first is true it does not follow that the latter is true as well. This is not the case in chains of conjunct clauses or causally related clauses: if the proposition "A and B and C" is true, it follows that "A and C" is true as well, and if "A leads to B leads to C" is true, it follows that "A leads to C"

[99] This is the interpretation O'Grady et al. (2015) seem to suggest on the basis of a different example they discuss in their paper. I think it can be argued for, however, I doubt whether this is the most natural analysis, given that it asserts the possibility of inserting the entire (recursively formed) noun phrase elsewhere in the sentence: [It was after a walk in the park, [when Nick was tipsy, [in the afternoon, [on November 22nd, [before Thanksgiving]]]], that Nick and Michelle started dating]. This might yield a "grammatical" sentence in the strict sense, but certainly not one that language users would easily produce in practice. Alternatively, one could argue that the relevant part of (8) is not one recursively formed noun phrase, but a string of serially combined noun phrases: [after a walk in the park,] [when Nick was tipsy,] [in the afternoon,] [on November 22nd,] [before Thanksgiving]. Following this analysis, which I think is the more credible from a language usage point of view, the sentence would be more like an elliptic version of a "narrative" presentation in multiple sentences: Nick and Michelle started dating after a walk in the park. Nick was a bit tipsy. It was in the afternoon on November 22nd, before Thanksgiving... See also Verhagen's (2010) discussion of "tail versus true recursion".

is true. If one of the elements in a string of conjunct or causally connected clauses does not fit with the story, the entire statement is false. However, if the intentional clause "B *doesn't know* about X" does not fit with the story, it does not follow that "A *thinks* that B *doesn't know* about X" does not fit with it. In other words, all that participants need to know for picking the correct option in the case of (8) is the answer to any one of the following questions: did the couple start dating before or after Thanksgiving? Was it on the 22nd or on the 29th? Was it in the morning or in the afternoon? In the case of (7), though, they *do* need to know something about what Victor knows that Amy knows that Sheila intends that John thinks that she thinks about him liking her. (However, note that participants may be crucially "aided" by the forced-choice design here: in fact, they only need to know whether Victor *does* or *does not* know about all this. More about this issue in the next section).

All in all, I think that question (8) embodies an unsuccessful attempt of the authors to solve all three of their own objections against the old test: taking a short-cut is possible after all, the question exhibits a different form of conceptual embedding compared to intentionality questions, and it exhibits a different form of syntactic embedding (or, arguably, no syntactic embedding at all, see note 99). A total of 14 out of 28 control questions in the relevant condition ("explicit"; see below) of O'Grady's study are structurally similar to (8) (cf. the supplementary material of O'Grady et al., 2015).

The other 14 questions in the relevant condition take the form of the one cited in (9). Do they do a better job of eliminating those objections? Each single clause is dependent on the clause to its left: for example, Sheila's supervisor's boss' friend is not the same person as Sheila's friend or Sheila's boss' friend. However, there is also a difference. It can be the case that John thinks that it is sunny outside, while Mary thinks that John thinks that it is raining. Yet it is not possible that John's mother is Mary's boss, while at the same time Mary's boss is not John's mother. In other words, Dennett's substitution test, explained in Chapter 1, Section 1.2.1, does not work for intentionality statements, but it does work for possessive relationships. As a consequence, spotting the one element that is at odds with the story can here also be done without processing the entire string: in the case of (9) all one needs to know is that Mike is Sheila's supervisor and not her best friend in order to choose the correct option: A.

Intentionality questions like (7) and memory questions like (9) may thus be matched in the sense that the concepts and the grammar exhibit recursive embedding[100] (the second and third objections are eliminated), but the possibility to short-cut processing keeps haunting these questions.

In the third study I have introduced at the beginning of Section 6.2, carried out by Haddad, the two types of questions are also matched more closely compared to the original questionnaire. Here the philosophy was not to introduce conceptual and syntactic embedding in the memory questions, but to concentrate on matching sentence length. The questionnaire as a whole is indeed well balanced in this respect (see the Online Appendix as referred to in note 95). In addition, broken-chain intentionality questions were eliminated. However, in this study it is also still possible to use processing short-cuts in the ways described in this section for quite a few of the questions.

Whether all of this is a problem or not depends on what is expected from the control questions: in, for example, Kinderman et al. (1998) factual questions serve to determine whether memory for details from the story is a factor limiting performance. However, given that the intentionality questions are different in a variety of respects, it may be problematic to make more precise comparisons between these questions (as has been done in, for example, Powell et al., 2010, where factual questions have been used as a baseline task). In some studies this may have led to interpretation problems, given that it is hard to determine to which part of the difference between factual and intentionality questions the findings have to be attributed. Although this problem is also noticed by O'Grady et al. (2015), their attempt to match the factual questions closer to the intentionality question comes with new problems, as I have discussed above. The same is true for Haddad's improved control questions.

[100] Interestingly, from the perspective of language usage, the resulting statements are also equally idiosyncratic: for both (7) and (9) it is impossible to think of a real-life context where such sentences would be uttered (except, perhaps, a humorous context like the one in the Friends episode discussed in the Introduction of this thesis).

6.2.4 True and false statements

Answering a question without processing it as a whole, i.e. taking a short-cut, is structurally more likely to be possible in questions where the right answer is "false". Consider the following factual memory questions taken from Haddad:

(10) Sam asked about finding a Post Office so that he could send a birthday present [false]
(question 13 in Haddad's questionnaire, see note 95)

(11) Sam couldn't send the card because when he got to the Post Office, it was closed [true]
(question 15 in Haddad's questionnaire, see note 95)

Haddad's version of the story about the Post Office is slightly different: the office prankster Henry goes by the name Helen and instead of wanting to buy a tax disc, Sam wants to send a birthday card to his grandmother. All that participants need to know to determine that (10) is false is that Sam was going to send a card, not a present, but to see that (11) is true, they have to know that Sam found the Post Office closed *and* that he was going to send a card. In other words, in "false" questions spotting *one* element that does not fit with the story is enough, whereas in "true" questions participants have to determine that *all* elements fit with the story. The higher the order of complexity of the question, the more this misbalance is amplified: in a second-order question it is one false element against two correct ones, but in a fifth-order question this ratio is one to five.

This issue is not limited to factual memory questions. Consider the following example:

(12) A. Megan wants Lauren to know that she, Megan, knows that Stephen knows that Elaine knows that Bernard feels she doesn't know him well enough to date, so that Lauren asks Stephen out [correct]

> B. Megan doesn't want Lauren to know that she, Megan, knows that Stephen knows that Elaine knows that Bernard feels she doesn't know him well enough to date, so that Lauren doesn't ask Stephen out [false]
> (see supplementary material of O'Grady et al., 2015)

Question (12) can be answered by knowing the answer to the simple (second-order) question: does Megan want Lauren to go out with Stephen or not? The clauses containing the "conclusions" ("so that Lauren asks Stephen out" and "so that Lauren doesn't ask Stephen out") have the same effect as had the broken chains discussed above: by inserting one causal link into the string of embedded clauses, it became possible to process the "cause" and the "consequence" as separate chunks. If one of the two did not match the story, the entire proposition was false. Similarly, in (12) it is possible to process the premise (involving eight orders of intentionality) and the conclusion (involving two orders) as separate chunks. Given the forced-choice design, plus the fact that the two conclusions present opposing scenarios, it follows that one of the conclusions *has* to be inconsistent with the story. What is an "eighth-order" question according to the authors, can in this case be answered by simply comparing two second-order propositions.

From this follows a crucial difference between the true/false design used in Brown and Haddad and the forced-choice design used in O'Grady: in the former it should be structurally harder to answer questions where the answer is "true", whereas in the latter this effect does not occur.[101] After all, in a forced-choice design it is *always* possible to find the right answer by spotting a single false bell. Whether or not this is an advantage can be debated: it certainly does increase consistency within the questionnaire, however, if the overall aim is to test participants on their ability to handle questions at different orders of complexity, it may be rather disruptive, as became clear with question (12).

Note that there are also false statements in which spotting the inconsistent element is possible only by processing the statement as a whole.

[101] In studies using a true/false design it may be advisable to calculate mentalising scores of individual participants on the basis of questions where the answer is "true" only. This was done by Van Duijn et al. (2014) in their study of mentalising performance in relation to school grades and personality traits: the questionnaire featured both true and false questions, but only the true ones were used to calculate the scores.

This goes for statements that exhibit an unbroken chain of embedding and, at the same time, do not involve unfamiliar or implausible elements. Consider the following examples:

(13) Pete thought that Helen wanted Sam to know that she realised that the Post Office was no longer on Elm St. [false]
(question 7 in Haddad's questionnaire, see note 95)

(14) A. Megan knows that Stephen doesn't know that Elaine knows that Bernard feels that she doesn't know him well enough to date [false]
B. Megan knows that Stephen knows that Elaine knows that Bernard feels that she doesn't know him well enough to date [correct]
(see supplementary material of O'Grady et al., 2015)

Both questions consist of statements exhibiting unbroken chains of embedding. On top of this, (13) refers to a scenario that is false, but that does have some credibility, given that the entire confusion in the story is exactly about Helen (alias Henry in Brown's version) intending or not intending to send Sam to the wrong street. Likewise, both options in (14) are credible, given that the story is precisely about Megan finding out what Stephen does or does not know about Bernard and Elaine.

The observations discussed in this subsection can be summarised in the issues (vi) and (vii) as follows:

issue (vi) Some questions exhibit a disproportionate increase in complexity per level for questions where the correct answer is "true" versus those where the correct answer is "false": if the correct answer is "true" participants have to check *every* element for consistency with the story, whereas questions where the correct answer is "false" can be answered by spotting *one* element that does not fit. This issue seems to be best avoided in questions that exhibit an unbroken chain of embedding and that present

a scenario that is (about) equally credible when thought of as true or false.

issue (vii) Given issue (vi), there is a crucial difference between questionnaires using a true/false design and those using a forced-choice design: when using the latter, spotting a single "false bell" is possible in *all* questions (although it is more difficult in questions exhibiting an unbroken chain of embedding and presenting two equally credible answer options).

6.2.5 Structure of interaction

As mentioned in the brief description of the study at the beginning of Section 6.2, the central aim of O'Grady et al. is to increase ecological validity by introducing movie clips that feature acted-out versions of the stories and questions. Whereas I make a distinction between narrative and propositional presentation in this thesis, they make a distinction between "explicit" and "implicit" presentation. Confusingly, at least in this context, is that by "explicit" they mean *both* the narrated stories and the propositional questions as used in the classic mentalising tests (which I have argued to be very different in kind in Chapter 2), and by "implicit" they mean their novel acted-out stories and questions. In four conditions they cross narrated stories with propositional questions (explicit-explicit), narrated stories with acted-out questions (explicit-implicit), acted-out stories with propositional questions (implicit-explicit), and acted-out stories with acted-out questions (implicit-implicit). The example questions cited in (7), (8), and (9) above are all from the propositional/explicit condition. The scripts for their acted-out counterparts from the implicit conditions are as follows (cited again from the supplementary material of O'Grady et al., 2015):

(15) A. Sheila: Anyway, so I chatted to Amy about it at the office and she reckons it's a good plan to just keep letting John think I haven't figured it out.
Victor: Yeah, Amy came by and told me about it all.
B. Sheila: Anyway, so I chatted to Amy about it at the office and she reckons it's a good plan to just keep letting John think I haven't figured it out.
Victor: That's weird, I spoke to Amy today. I had no idea that she knew about this situation.

(16) A. Nick: We started dating before Thanksgiving, in the afternoon on November 22nd. I was a bit tipsy and we'd just got back from a walk in the park.
B. Nick: We started dating after Thanksgiving, in the morning on November 29th, while we were walking in the park. I was a bit tipsy at the time.

(17) A. Amy: Yeah, it's really complicated. So your best friend is Sheila's supervisor Mike's boss's PA, John, and Shaun is my brother.
B. Amy: Yeah, it's really complicated. My boyfriend is PA to Sheila's best friend Mike's supervisor's boss, John, and Shaun is my brother.

An important innovation here according to the authors is that they, by introducing their acted-out presentation form, have managed to present *conceptually* embedded information without using embedded sentences. Note that this was already done in the narratives used in the original tests. However, a novelty indeed is that they also have two conditions in which the *questions* are presented without using embedded sentences.

It is indeed true that (16) and (17) look a lot closer to normal language usage than their propositional/classic counterparts (7), (8). However, it should be noted that (17) retains the problem I have pointed out above: it is still possible to find the right answer just by knowing whether it was before or after Thanksgiving, on the 22nd or the 29th, or in the morning or afternoon—each of

which is literally given in the story that the participants have just heard or seen in acted-out form. Question (18) is only minimally disentangled compared to the embedded-sentence version in (9). Yet if anything, this has made it even easier to spot the element that does not fit the story. The same goes for the acted-out/implicit version of questions (12) (boldface added):

> (18) 7. A. Megan: So, I'm thinking that Lauren needs to know what I heard. Right? Because if she knows what I know right now, about Elaine's crush, and Bernard's rejection, and that Stephen knows the whole thing…she'll work up the guts to ask him out! **So I'm going to tell her tonight**.
> B. Megan: Well, if you think about, if Lauren knew what I heard today – and if she knew that Stephen knew all about it too, about Elaine's crush and Bernard's weird reason for rejection and everything – she'd ask Stephen out. But I don't want her to do that, **so I'm not going to tell her**.

As pointed out above, it was possible with (12) to bypass the processing of the eight-order string of embedded intentional states. This is made even easier in (19), especially in the implicit-implicit condition, given that the acted-out/implicit version of the story ends as follows (boldface added):

> (19) Megan: Right! So, Lauren doesn't want to ask Stephen out because she thinks he's into Elaine – but if she knew that Stephen knows that Elaine likes Bernard, and that Stephen knows that Elaine's not into him, she might work up the guts to ask Stephen out.
> Chris: I guess…so are you going to tell her?
> Megan: **Yeah, I'm going to tell her the whole thing tonight**.

All participants need to remember to answer the question that allegedly embodies the highest level of complexity in the test, are the ten final words of the acted-out story.

O'Grady et al. (2015) present as their central finding that participants performed strikingly well at all levels of complexity, especially in their implicit-implicit condition. Performance did not drop drastically at any level, as was

claimed on the basis of the classic tests. They argue that this is probably due to the high ecological validity of their stimuli: according to them our natural human social ecology is full of higher-order intentional processing tasks. By designing stimuli that mimic this ecology as closely as possible, they claim to have shown that participants can almost effortlessly handle tasks up to eight orders of intentionality. However, I think that their test is highly ecologically valid precisely because participants do *not* need to process long strings of embedded intentional states, as shown in my analysis, but instead can rely on simple cues that bypass such processing when deciding between alternative scenarios. Other than in the classic tests, where participants are at least sometimes forced to deal with idiosyncratic statements that have to be processed as single units, O'Grady et al. (2015) allow participants to be normal mindreaders—and that is: lazy mindreaders.

6.3 Concluding remarks

In this chapter I have distinguished five central conclusions brought forward by the mentalising paradigm. In short:

- error rates of intentionality questions show a steep increase at complexity level 6, suggesting a limit to the ability of participants to handle embedded intentional states;
- between-subject variation suggests that some individuals have their limit around level 4, others around level 5, and again others around level 6 or even higher;
- although scores on intentionality questions and memory-control questions tend to be related, the steep increase in error rates at level 6 cannot be explained in terms of memory performance only, suggesting that there is something especially challenging about reasoning with embedded intentional states;
- participants' mentalising scores correlate with other measures of these individuals' social aptitude;

- participants' mentalising scores correlate with the amount of grey matter in relevant brain areas.

These conclusions are based on a series of studies, in which the general pattern has been replicated multiple times. However, nearly all of these studies have used versions of the same questionnaire, featuring stories describing social events, followed by forced-choice or true/false questions with embedded intentional states and factual details. In this chapter I have discussed issues connected with this way of testing mentalising competence, pertaining to the discrepancy between a narrative presentation in the stories and use of embedded sentences in the questions, inconsistencies regarding viewpoint layers that are sometimes "unpacked" and sometimes left implicit, structural differences between questions exhibiting embedded clauses and those with clauses that are related in different ways (e.g. conjunct or causally related), structural differences between true and false questions, and the gap between the use of intentional reasoning in the test and in real-life interaction.

The recent study by O'Grady et al. (2015) has raised doubt regarding the conclusions from the mentalising paradigm, partly based on the same issues with the questionnaire that I have pointed out here. However, as I have also discussed in this chapter, their own updated mentalising test, while having promising aspects, is also partly haunted by some of the old issues and for another part comes with new problems. Therefore, more research is needed before anything reliable can be said about the consequences for the central conclusions of the paradigm.

For now, I think, there is another important puzzle still unsolved: how can it be explained that the mentalising tests, despite all issues, produce meaningful variation correlating with measures of participants' social lives and overall aptitude in the social domain? I see two non-exclusive explanations:

- There are problems with quite a few of the questions, but others (such as (13) and (14) above) work well in the sense that they genuinely force participants to process the entire task they pose as one single unit. The meaningful variation in mentalising scores between participants could be principally driven by these questions. I have argued throughout this thesis that the processing of such unbroken chains of embedded intentional states is something we do not do by default when interacting,

but rather in exceptional cases, such as anticipating or repairing a misunderstanding. This can explain correlations with measures of social aptitude: if filling out a mentalising test is like handling exceptional situations in social interaction, those participants who are better at the test are also better at trouble-shooting whenever an interaction threatens to break down—a skill that may well be a good indicator of general social aptitude.

- The way in which the questions, and to some extent also the stories, present the fictive social situations used in the mentalising tests may be unnatural and problematic, however, the test itself creates a new and real situation of social interaction: the one between experimenter and participant. The experimenter has designed the questions and decided what the correct answers are. Some participants may be better at estimating what the experimenter wanted them to do, and be more motivated to figure this out in the first place. They may pick up even the smallest cues (like those in questions 7 and 10 from Brown's study, see also note 94 above) directing them towards the answer that the experimenter had in mind. This is a point that may theoretically produce biases in any test involving questionnaires, but in this particular case the bias happens to overlap with the target variable: being better at estimating the experimenter's intentions may indicate greater general social aptitude.

Both these explanations thus suggest that the associations between mentalising scores and other factors relevant to participants' social lives were not produced because the tests accurately "mirrored" the complexity generally involved in human interaction, but because they assessed participants on factors indicative of their ability to deal with special (partly extreme) cases of such interaction. This does not mean that these associations are no longer meaningful, but it does shed a different light on the foundations of social cognition: although it may be possible to assess general social aptitude using a task that forces participants to deal with embedded intentional states, this task should not be used as a model for what people do on a cognitive level in everyday social contexts.

Conclusion

In the Introduction I have put forward a series of observations and questions, and set out as the overall aim for this thesis to work these out in detail, thereby rethinking the nature of the complexity posed by intentional states underlying human interaction events. Throughout the subsequent chapters, I have looked at how such complexity should be thought of *conceptually* and how it can be handled *linguistically*. Also, I have made suggestions about how human agents are able to process it *cognitively*. I have pointed out that this thesis' methods and frameworks are ultimately rooted in the humanities, but that its topic requires analysing and bringing together insights and material from areas across both the humanities and sciences, in particular linguistics, literary studies, philosophy, and various disciplines within the social, cognitive and biological sciences. Therefore I have characterised the overall project as being an example of *topic-oriented scholarship*: instead of working with the questions, assumptions, and methods common in one academic discipline, I have applied a pragmatic mix of expertise and methodology I considered suitable for making progress on the chosen topic.

The observations and questions put forward in the Introduction are grouped in thematic rubrics: complexity posed by multiple intentional states, representation of complex thoughtscapes in discourse, communicative and cognitive economy, and the consequences for existing theories and frameworks, in particular our evolutionary story. Given that Chapters 2-6 already end with a section summing up discussion points and conclusions, I will structure what follows using these rubrics rather than going through the final remarks in the "chronological" order of the chapters. However, before doing so I will first provide an overview that will pull together the lines connecting this thesis' main findings and insights.

Overview

Human interaction is characterised by an endless "polyphony", a perpetuated multitude of perspectives and perspectives-on-perspectives. Not only do we share and coordinate our own inner life with that of the people we interact with, but we also constantly make implicit and explicit reference to the intentional states of others who may or may not be present at the time of speaking, or who may even exist only in the imagined worlds of thought and fiction.

In philosophy of mind and the cognitive sciences, this polyphony of perspectives has often been conceptualised as a series of *embedded layers* of the form "A *thinks* that B *understands* that C *expects*...etc.". Building on this tradition, tests have been devised targeting what is referred to as "multiple-order intentionality" or "mentalising", generally finding that humans can handle up to around five layers of embedded complexity. This idea was subsequently implemented in theories and frameworks concerning such topics as cooperation, moral reasoning, social aptitude and variations in quality and size of individual social networks, (a)typical development, language competence, and appreciation and production of literature.

However, throughout this thesis I have suggested that the conceptualisation of mentalising involving series of embedded layers stands in stark contrast to how dealing with a polyphony of intentional states takes shape in actual discourse and interaction. Firstly, it appears that intentional states are normally connected and interlinked in all kinds of different ways, forming what I have termed a "thoughtscape" rather than a recursive string. Secondly, if a (complex) thoughtscape is being represented in natural discourse, the labour of indicating the connections between intentional states is generally distributed over a wide variety of linguistic elements across lexical, grammatical, and narratological categories. Hence, representing a thoughtscape by a proposition featuring only recursive embedding of clauses frequently yields an infelicitous rendering of the actual relationships between the intentional states. Thirdly, parts of the thoughtscapes underlying interaction events come "packaged", and they are unpacked only if the context so requires. Fourthly, instead of thinking of human interaction as a process in which individuals seek to "join" their

intentional states through working out what others *intend* one to *understand* that they *want* one to *believe*, and what one *intends* others to *think* that one *intends* them to *believe*, such interaction should rather be seen as a joint activity in which cues are provided negotiating how a set of already shared beliefs (referred to as the common ground following Clark, 1996) should be updated. Fifthly, linguistic items, ranging from lexical units to grammatical patterns and narratological structures, embody a wealth of experience accumulated from generations of language users attempting to coordinate their own intentional states with those of interlocutors and third parties, regarding the non-intentional world by which they are surrounded. As such, linguistic items can be thought of as "supercues" supporting the process of negotiating how a thoughtscape should be conceptualised: depending on context and signaller's goals, language can offer precise analytical tools for working out details and nuances of how intentional states are mutually connected, or, conversely, language can offer cues referring to entire chunks of a thoughtscape holistically, leaving such details and nuances packaged in order to serve convenience and efficiency. In this way, the "toolkit" offered by the symbols of a language contributes at once to the richness and detail of the total system, as well as to its economy and workability. Sixthly and finally, by acquiring the lexical, grammatical, and narratological conventions that constitute this toolkit children become full-blooded human mindreaders—who are as a rule *lazy* mindreaders: operating in a socio-cultural environment that contains the coagulated interactional experience of earlier generations, saves processing costs on the individual level.

Complexity posed by multiple intentional states

As an alternative for conceiving of intentional-state complexity as a series of embedded layers, I have suggested the notion of the *thoughtscape*, defined in Chapter 1 as the total network of interlinked intentional states that are in some way relevant in the course of an interaction event. With respect to this, there is a question brought up in the Introduction that has not yet been answered: can the concept of the thoughtscape function as an alternative measure of intentional-state complexity? In this section I will suggest that this is possible,

but that an accurate implementation of the thoughtscape as a measuring tool requires further investigation of the qualitative differences between different sorts of intentional relationships.

The idea of embedded layers clearly comes with a natural complexity scale: counting layers and comparing numbers of layers across tasks, individuals, species, and so on, seems to be inherent in the very concept of multiple-order intentionality. However, I have pointed out that this is problematic: when the orders of intentionality are used as a meta-linguistic measuring tool to form paraphrases of how intentional states underlying an interaction event are related, they often force these intentional states into a strait-jacket that misrepresents the actual complexity. Paraphrases featuring only embedded clauses are at the same time too complex *and* too simple: on the one hand, they easily produce opaque renderings of a situation that is not too hard to grasp as such or in the form of a story, while, on the other hand, such propositions often fail to convey all kinds of nuances and details. A good example is Zunshine's paraphrase discussed in Chapter 3, which is indeed much harder to understand than the relevant passage of the novel itself is, while at the same time underrepresenting the subtle perspective shifts characteristic of the 360-degree view offered by Woolf's prose. Even in the case of *Othello* this holds true: as I have argued in Chapters 2 and 5, this play has the quite exceptional feature of meaningfully embedding a relatively large number of perspective layers. However, this only works because its plot combines multiple scenarios that come with significant *differences* in shared knowledge or common ground (scheming plans, revenge, adultery), and, as argued in Chapter 5, it is precisely in those cases that it is relevant to work out what A *knows* that B *intends* that C *thinks*...(etc.) In that sense, the embedded proposition "Iago *intends* that Cassio *believes* that Desdemona *intends* that Othello *considers* that Cassio *did not intend*..." *does* in some way accurately summarise the situation with which the audience is confronted after the first two Acts. Even so, this proposition suggests a sense of opaque complexity that is unlikely to be perceived when following the play's narrative on the one hand, while being far too narrow to catch the full complexity of the thoughtscape that has emerged by the end of Act II on the other hand.

In the case of both *Mrs Dalloway* and *Othello* intentional-state complexity was better conceptualised as a thoughtscape than using a proposition featuring a string of embeddings. However, although the relevant parts of both texts were misrepresented by such a proposition (that is, the first more so than the latter), it offered an easy way to *compare* the two in terms of the assumed intentional-state complexity, simply by counting the number of embeddings. Such counting is much less straight-forward when comparing two thoughtscapes. To recapitulate, in Chapter 1 and 5 I have discussed that the basic unit of a thoughtscape is the interaction event, in which two or more parties interact using linguistic and/or non-verbal cues. Prototypically, these are interlocutors in a face-to-face setting, but in the special case of fiction the position of the speaker is taken by the narrator and that of the addressee by the reader, hearer, or spectator. Speaker and addressee have a common ground, a set of shared beliefs, which they update in the course of the interaction event. This does not by default include reasoning about intentional states, but both the intentional states of the speaker and addressee and those of third parties can become relevant depending on setting and context (whether they are subsequently spelled out in language or left implicit is another question—see the next section). For example, when a narrator tells a story about two characters, it makes sense to use linguistic items that invite the reader to form representations of these characters' intentional states at various points in the development of the plot. These intentional states are both embedded in the perspective[102] of the narrator, but not necessarily into one another. They can exhibit causal relations (A *thought* X, which made B *want* Y), form meaningful conjunctions from the perspective of the reader (A *thought* X and B *thought* Y), or, indeed, be embedded (A *believes* that B *thinks* that X). A thoughtscape is therefore, as I have argued, best conceptualised as a *network* of mutually linked intentional states. This network emerges in the course of an interaction event to the extent that particular contexts require interlocutors to draw inferences about each others' or third parties' intentional states. For example, the context

[102] At the beginning of Chapter 4 I have made the distinction between "intentional state/mindstate" and "perspective/viewpoint" roughly as follows: intentional states or mindstates are atomic "snapshots" of a subject's relation to an object; a perspective or viewpoint comprises the broader total of an actor's subjectivity of which intentional states are isolated parts or "snapshots".

of a practical joke may require A to make B believe that X is the case, while at the same time telling C that B thinks Y; or, in the context of a misunderstanding A may wonder what B intended A to think when saying X.

With this in mind we can turn to the question of measuring intentional-state complexity. Having the form of a network, thoughtscapes can in principle be used for this purpose: counting the number of intentional agents and relationships between them (i.e. the number of edges in a network graph) can work as a global *quantitative* measure of thoughtscape complexity. However, the practical implementation of this measure is not straightforward, given that one would also have to account for the fact that the relationships (i.e. edges) can be *qualitatively* different in complexity: "A *thinks* that X and B *thinks* that Y" may be easier to process than "A *suspects* that X and B *does not think* that Y" or "A *wants* B to *believe* that X". Getting a grasp on such qualitative differences can begin by incorporating existing psycholinguistic insights into processing grammatical procedures such as negation or anaphoric reference into the framework of the thoughtscape. In this way, single relationships can be compared and weighted for the amount of processing effort they require. On top of that, it will be necessary to develop new research that evaluates the processing implications of incrementally integrating multiple intentional relationships. With the "old" conceptualisation of complexity as embedded layers it seemed evident that every added layer led to an increase in cognitive load, until a limit was reached at around five orders (or, according to some authors, a higher number—see Chapter 6 for a discussion). Contrastingly, it is not clear whether every additional edge in a network of intentional states should be seen as a factor adding cognitive load, and whether this adds up to a maximum network size in terms of what individuals can on average process cognitively. In Chapter 2 and 3 I have suggested that it is not plainly the case that cognitive factors limit story complexity, but rather that understanding reaches as far as the maximum complexity that can be covered by a story. At a glance this may seem a sophistic twist instead of a real claim, but think of it this way: if the reader's, hearer's, or spectator's understanding of a network of intentional relationships depends for an important part on factors inherent in the language and narrative used, it must be assumed that stories employing a more optimised mix of such factors can represent more complex networks. In this way, the limits of what is the *most*

complex network that can be covered by a story can be pushed until, at some point, the story "breaks down" in terms of coherence, becomes unreadable, or in another way fails to do the job of exposing its plot to its addressees. In this view, the upper boundary to handling intentional state complexity is not *given* by individual cognitive limits that have an average height in human populations, but instead dynamically *produced* by an interplay between individual cognitive factors and group-level cultural phenomena of language and narrative. Such a view has consequences for the evolutionary story set out in Chapter 1 and fits with two other pieces of the puzzle: one pertaining to linguistic symbols as "supercues" and the other to cognitive economy.

Representation of thoughtscapes in discourse

Far from all intentional states relevant to interaction events are highlighted or made explicit. Consider once again the picture with the woman on a staircase carrying a suitcase in Figure 1 of Chapter 1 (it can be debated whether standing in the position of the photographer and deciding to offer a helping hand counts as "interaction", but let us for the sake of the argument assume that it does). Within this interaction event as seen from the perspective of the person standing in the position of the photographer (referred to as the *mindreader* in Chapter 1), there is a role for the intentional state of the woman with the suitcase: an estimation of this intentional state (the *mindread*) is one of the factors on the basis of which the mindreader decides whether to take action.[103] It is possible to formulate an explicit paraphrase of the mindread (e.g. "mindreader *thinks* that mindreadee *intends* to carry the suitcase upstairs"), but in most cases the mindread will remain an implicit factor in a decision about future behaviour, which can be non-linguistic (provide a helping hand) or linguistic (e.g. "Can I help you?"). Only in exceptional cases will an utterance in the practice of everyday interaction come close to an explicit paraphrase. In this

[103] Later on in Chapter 1 I have nuanced this view by suggesting that this can be safely said on the W-level of what the task is, but that caution is needed on the H-level of how the task is carried out: in this context, stating that the intentional state of the person on the staircase is a factor on the W-level does not entail that the process "running" in the mindreader's mind necessarily resembles that in other contexts in which mindstates play a role on the W-level (see Section 1.1.4).

example that could happen when the mindreader appears to have made the wrong estimation after all: imagine that the mindreadee wanted to test whether the suitcase would be too heavy to be carried around for a week. In that context, a conversation could be imagined in which she declined help, after which the mindreader said "I *thought* you *intended* to carry it upstairs…"

I have argued that interaction events should be seen as a negotiation process between speaker and addressee of how a (presumed) set of shared beliefs or common ground should be updated. Whereas it is in principle possible to carry out such negotiation without language, language comes with powerful, specialised cues (or *tools*) on several levels (lexicon, grammar, narrative) to facilitate this process: from packages that project entire scenarios at once (e.g. "Sorry, I *misunderstood*") to all kinds of tools for indicating subtle perspective shifts, epistemic stances, and so on, to work out the details (e.g. "I *see*! You *wanted to try* whether…but to me it *looked* as if…I just *wanted* to be helpful…"). It is significant to notice that a context in which the help is appreciated goes much more naturally without words than one in which help is declined. In the first case, both parties are "on the same page", whereas in the latter case there is clearly a difference in how they envisage the interaction to continue—this once again suggests that contexts with *differences* in common ground are the ones in which working out who-thinks-what is relevant.

In Chapter 5 I have argued that the dynamics of negotiations about how to update the common ground can be pictured as having three dimensions, with cues serving coordination between the intentional states of signaller and addressee ((*x*)-axis), with respect to objects of joint attention ((*y*)-axis), and with respect to third-party perspectives ((*z*)-axis). From a "synchronic" perspective looking at one single interaction event, each linguistic item supports this process because both interlocutors share knowledge about how it operates on one or more of these axes (i.e. their *meaning* in this model). From a "diachronic" perspective, however, it can be said that every lexical item and grammatical procedure ultimately is the result of generations of language users trying to coordinate their mindstates in interaction with each other and the environment, thereby converging on solutions that are communicatively effective, cognitively efficient, and learnable for new generations of language users. In this sense, when we use language today, we have "supercues" at our

disposal that embody the accumulated interaction experience of generations of language users.

Evolution of the lazy mindreader

The advocated view allows individuals to be "lazy" in multiple ways. In Chapter 5 I have argued that Scott-Phillips (2015) and Sperber (2000) misconstrue the complexity of the mindreading abilities needed for present-day human interaction *in theory*: rather than working at several levels of intentionality by default, individuals make the a priori assumption that a set of knowledge states (the common ground) is mutually shared. In terms of investing processing effort, this allows them to start from zero and scale up if necessary.[104] On top of that, I have suggested various mechanisms and principles that *in practice* save individuals from having to scale up most of the time, including the observation from relevance theory and the study of alignment that interlocutors operate "in complementary predicaments" (Apperly, 2011: 115, referring to Sperber and Wilson, 2002), and the idea that most everyday interaction exhibits a structure of testing-adjusting-retesting in which representations have to be just "good enough" for the interaction to continue, but no better. These views were also consistent with observations made in Chapter 2 and 3, where the focus was on the exceptional situation posed by some works of (literary) fiction: rather than being challenged to the extreme by the necessity to keep track of complicated networks of intentional states in order to follow the plots, it appeared that the reader could "sit back and relax" while the narrator brought a rich mix of strategies to bear gradually exposing all the involved intentional states and their mutual relations.

A different part of the story is that human children can not only use linguistic elements they acquire for communication, but they presumably also extend their thinking repertoire through the process of learning how to use

[104] Note that this view also solves the "infinity paradox" mentioned briefly in the Introduction. Interlocutors normally do not have to "meet" the other's mind a few mindreading steps down the line (in which case there are always more steps possible—the infinite regress, as many authors suggest); instead, they are already "together" at zero by default. If each step is jointly taken from there (through communicative negotiation), there is no longer a reason why an infinite regress would be lurking.

some of them. It is important to note that this works for all individuals within a particular cultural-linguistic community *in the same way*: given that A and B are both (adult-level) members of the same community, individual A can not only presume a particular set of beliefs to be mutually shared with B, but she can also be sure that B has the same "toolkit" on board for negotiating how these beliefs should be updated or how differences in common ground should be detailed. Examples of tools for such negotiation (as discussed in Chapter 4 and mentioned in the former section) include viewpoint packages, supporting the possibility to project entire scenarios holistically, and various grammatical patterns making it possible to work out the details of who-knows-what one by one. If A uses the word "mislead" she at the same time activates in herself and in B the complete topology packaged by this word, including the roles of a misleading and a misled party and the default intentional relationships between them. Both interlocutors can benefit from this in the next steps of their interaction by integrating the topology with context-specific details.

This view also implicates a special niche for narratives of all kinds, ranging from the day's latest gossip to myths, parables, and even literature. Acquiring the tools for negotiating and coordinating intentional stances comes down to learning how they are used in different contexts. Stories of all kinds provide a rich and varied environment for such learning. Whereas "mislead" may perhaps occur frequently enough in everyday contexts for language learners to acquire its full topology, this may be different in the case of, for example, "scheming plan" and the scenario attached to it on which *Othello's* plot builds. I suggest that one of the functions of our tendency to tell stories is that it allows language users to get accustomed to a wide set of such tools available in their cultural-linguistic community.

The question finally remaining now, is how all of this affects our evolutionary story. It is important to emphasise the qualification I have put forward in Chapter 1 once again: the material studied throughout this thesis and the arguments developed on the basis of it should not be taken as an attempt to "prove" a particular chronology of evolutionary events. However, given that the ways in which different scenarios of human cognitive evolution can be compared are necessarily limited (archaeology and comparative biology can tell us only so much in this domain), I have suggested to look at how

complex thoughtscapes are handled in the actual practice of human interaction. What exactly do we find in a play or novel renowned for dealing with wide arrays of perspectives? How do newspapers report on cases in which multiple intentional states are relevant? What can we learn from usage-based linguistic approaches as developed by Clark (1996) and Verhagen (2005)? How do participants seem to deal with experimental tasks forcing them to consider multiple intentional states at the same time? By investigating such questions one by one, a view has emerged that certainly can be said to be *more compatible* with one evolutionary scenario than with the other.

In this light I have suggested a "move" that is conceptually comparable to the one made by Shultz, Opie, and Atkinson (2011):[105] the route to a cultural-linguistic community should not be seen as one in which individuals developed increasingly profound mindreading abilities, going step-by-step from "A intends that B believes…" to "A intends that B believes that A intends…" to "A intends that B believes that A intends that B understands…", and so on, until addressees and speakers had arrived at fourth- and fifth-order intentionality respectively, which they needed to establish the sort of cognitive "interdependence" Scott-Phillips (2015) and Sperber (2000) consider to be a prerequisite for human ("ostensive-inferential") communication. Instead, the alternative supported in this thesis is that a form of shared intentionality *preceded* the capacity to deal with complex thoughtscapes, including the ability to work out in detail what A believes that B believes that A believes (…etc.). Roughly, instead of interaction in dyads growing more sophisticated, gradually including more "minds" and thereby eventually forming the basis for cultural-linguistic communities, I suggest that the presence of such a community was needed for increasingly sophisticated interaction in dyads.

At the end of Section 5.2.3 in Chapter 5 I have discussed the cognitive structure individuals need for participating in a system of interaction based on shared intentionality from a *synchronic* perspective, thus abstracting from the *diachronic* story of how such a system could emerge over time (see also the first

[105] Clearly, there are also important differences: their model concerns going from a situation in which individuals form no group at all to one in which there *is* a social group, whereas my suggested move concerns a situation in which there already is a group to one in which individuals within this group form a cultural-linguistic community.

note in Chapter 5). I have listed the following abilities: (i) distinguishing between individuals and recognising whether or not they belong to our social network and/or cultural-linguistic community (in order to identify the appropriate sort of common ground); (ii) keeping track of former interactions and accessing "records" of this in real time; (iii) applying the principle of self-other equivalence; and (iv) assuming a particular set of intentional states to be shared with one's interlocutor and singling out individual perspectives only if the context requires to do so. In order to put together the diachronic story it would be necessary to determine when and to which degree these abilities were available to our ancestors. I think that more research efforts should be invested in this in the near future;[106] however, having said that, I will end by giving it a provisional shot.

The abilities (i) and (ii) are in principle primate skills, found in our great-ape relatives to a degree of sophistication that I could imagine to be sufficient for getting started. In my view there is thus no reason that these would have been factors limiting the emergence of the sort of interaction system suggested here at some point in our evolutionary path after the divide from the other great-ape lineages (that is not to say that we have not become better at these at later stages). Ability (iii), applying the principle of self-other equivalence, is what I have discussed as the skill needed for taking part in the "Schelling mirror world" (Levinson, 2006) in Chapter 5. It is fair to assume that this ability requires a degree of mindreading beyond what our great-ape relatives can do, but there is no reason why it should involve higher-order intentionality: interlocutors have to reflect on their own and the other's perspective, and they would need to be capable of imagining that roles be reversed in order to consider contributions to the interaction from both sides. This involves working with two related, but not necessarily embedded, intentional states in a flexible way.

Fitting ability (iv) into the diachronic story first involves specifying the conditions under which it makes sense at all for individuals to enter an

[106] As mentioned in Chapter 5, Stolk's (2014) experimental work provides an important basis, but it comes with the unavoidable issue that the participants have already been through the process of growing up in a present-day cultural-linguistic community.

interaction event under the assumption that a set of intentional states is mutually shared. I suggest that an attempt to deal with this issue should at least include the elements of *cooperation* and *repeated interactions* between a somewhat stable set of individuals. In this way, it is of interest to individual participants to find an efficient solution for achieving a common goal. Roughly put, if repeated interactions lack common goals, there is no incentive for individuals to invest effort in coming to any solution; if common goals have to be achieved only occasionally there is little incentive for coming to an efficient solution. It is when trying to achieve a joint goal in an efficient way that it makes sense to assume a set of shared intentional states.

This then forms a context in which the emergence of conventional signals for singling out individual perspectives and negotiating deviations from a common ground makes sense—after all, entering an interaction event under the assumption that a particular set of intentional states is shared may form a good *starting point*, but if the exact same set is also necessarily the *end point* of the interaction, it has little practical use. With the availability of signals for negotiating how this set of shared intentional states should be updated, such usefulness increases. In this way, I suggest that a co-evolutionary loop was triggered in which increasingly profound negotiation signals led to wider applicability of the assumption of shared intentionality and vice versa, eventually producing the cultural-linguistic communities in which the sophisticated linguistic items I have argued to be crucial for both communication and processing of complex thoughtscapes could coagulate and get passed on from generation to generation. As soon as such items are available to members of a group, they introduce a sort of supra-individual order capable of orchestrating joint actions, saving individuals lots of negotiation costs including time, risk, and cognitive power.

In Chapter 1, Section 1.3.2, I have discussed the "bonding gap", the idea that our hominid ancestors needed to find new ways of bonding increasingly large social groups. Whereas our primate relatives, and presumably also the last common ancestors we share with them, use(d) time-consuming one-on-one activities (social grooming in particular), our ancestors must at some point have relied on more efficient alternatives such as laughing, dancing, and singing together. It has been suggested that language also played a key role, but

contributed to social bonding in a more indirect way through, for instance, gossiping or sharing jokes and stories. These activities (except maybe telling basic jokes) have been argued to require the availability of a fairly sophisticated form of language, and given the assumption that such a language demanded around five orders of intentionality to be in place, its arrival was positioned late in our evolutionary history. Although I agree that storytelling involves quite sophisticated language forms, I have argued that higher-order intentionality is not a prerequisite for this, and that the ability to deal with complex thoughtscapes is not so much inherent in individual cognitive capacities, but rather produced by an interplay between individual cognitive factors and phenomena that are part of language and narrative itself. This points towards a scenario in which linguistic and narrative abilities, cultural-linguistic communities, and increasingly sophisticated mindreading skills have co-evolved rather than one in which the latter are a prerequisite for the former to emerge.

To conclude, the key elements in my provisional version of the diachronic story of interaction based on shared intentionality that are *not* likely to have been present in the last common ancestor with our great-ape relatives, are thus the mindreading skills for operating in the Schelling mirror world and forms of cooperation among somewhat stable groups of individuals geared towards achieving common goals. Gamble, Gowlett, and Dunbar (2014: 140-146) convincingly argue on the basis of brain size, complexity of artefacts, and patterns of their socially mediated activities on landscapes that these elements were available to the makers of the famous Acheulean hand axes: *homo erectus* who lived in the Lower Palaeolithic from around 1.7 million years ago. Given my arguments in this final section there is clearly no reason to assume that our ancestors at that stage possessed fully-blown linguistic abilities, instantaneously involving such sophisticated features as gossiping and storytelling. However, I suggest that the capacity to deal with higher-order intentionality can be no reason for why the co-evolutionary loop described above could not have started in this era. From there, I suggest that our social brain includes an *individual* and a *collective, socio-cultural* dimension. Our evolution in increasingly complex social environments has not just put pressure on our individual brains to get bigger and more powerful (i.e. the classic social

brain hypothesis), but also on finding culturally transferable solutions for using these brains in an optimally efficient way. These solutions can today be found in the toolkit that language and narrative offer us for dealing with complex social situations and their underlying thoughtscapes.

I very much look forward to future arguments and evidence supporting or challenging my version of our story, thereby shedding more light on the process of which the outcome has been studied in this thesis: on the individual level we are lazy mindreaders who could do little intentional reasoning with our "bare" brains, but empowered with the toolkits available in our cultural-linguistic communities we appreciate Shakespeare and Woolf, project complex thoughtscapes when reading headlines, enjoy *Friends*-style practical jokes, or can even consider taking up a career as a double agent.

References

Apperly, I. (2011). Mindreaders. The Cognitive Basis for Theory of Mind. New York: Psychology Press.

Apperly, I. and S. Butterfill. (2009). 'Do humans have two systems to track beliefs and belief-like states?' *Psychological Review* 116: (4): 953–970.

Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. London and New York: Routledge.

Austin, J.L. (1955). *How to Do Things with Words*. Oxford: Oxford UP.

Austin, J.L. (1962). *How To Do Things With Words*. London: Oxford UP.

Baillargeon, R., R.M. Scott, Z. He. (2010). 'False-belief understanding in infants.' *Trends in Cognitive Sciences* 14: (3): 110–118.

Bakhtin, M.M. (1984). *Problems of Dostoevski's poetics*. Minneapolis: University of Minnesota Press.

Bakhtin, M.M. (1981). *The Dialogic Imagination*. [Edited by M. Holquistt. Translated by Caryl Emerson and Michael Holquist]. Austin/London: U of Texas P.

Bal, M. Narratology. (2009). *Introduction to the Theory of Narrative.* Toronto: Toronto UP.

Barkow, J.H., L. Cosmides, and J. Tooby (ed.). (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford/New York: Oxford UP.

Baron-Cohen, S. and S. Wheelwright. (2001). 'The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism.' *Journal of Child Psychology and Psychiatry* 42:(2): 241–251.

Baron-Cohen, S., H. Tager-Flusberg, and M. Lombardo (eds.). (2013). *Understanding Other Minds*. Third Edition. Oxford: Oxford University Press.

Barthes, R. (1967). 'The Death of the Author.' *Aspen* 5-6: [item] 3.

Beekhuizen, B.F., and M.J. van Duijn. (2013). 'Packaged and Ready-to-Go: Linguistic Tools for Communication and Mental Processing.' 12th International Cognitive Linguistics Conference (ICLC), University of Alberta, Edmonton, Canada.

Behrens, T.E.J., L.T. Hunt, M.W. Woolrich, and M.F.S. Rushworth. (2008). 'Associative learning of social value.' *Nature* 456: 245–249.

Bennett, J. (1976). *Linguistic behaviour.* Cambridge University Press.

Bernaerts, L., D. De Geest, et al. (eds). (2013). *Stories and Minds: Cognitive Approaches to Literary Narrative*. Lincoln: University of Nebraska Press.

Bird, G. and E. Viding. (2014). 'The self to other model of empathy: Providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia.' *Neuroscience and Biobehavioral Reviews* 47: 520-532.

Boesch, C. (2005). 'Joint cooperative hunting among wild chimpanzees: Taking natural observations seriously.' *Behavioral and Brain Sciences* 28: (05): 692-693.

Boyd, B. (2009). *On the Origin of Stories. Evolution, Cognition and Fiction.* Cambridge MA/London: The Belknap Press of Harvard UP.

Breedveld-de Voogd, C.G. 2014. 'Gedachtenlezen in het burgerlijk recht'. Amsterdam: Amsterdam UP.

Brentano, F. (1995 [1874]). *Psychology from an Empirical Standpoint*. London: Routledge.

Budelmann, F. and P. Easterling. (2010). 'Reading minds in Greek tragedy.' *Greece and Rome* 57: (2): 289–303.

Burling, R. (2005). *The Talking Ape: How Language Evolved*. Oxford: Oxford UP.

Burton-Chellew, M. N., and West, S. A. (2012). Pseudocompetition among groups increases human cooperation in a public-goods game. *Animal Behaviour*, *84*(4), 947-952.

Byrne, R.W. and A. Whiten. (1988). *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*. Oxford: Oxford UP.

Byrne, R.W. and A. Whiten (eds.). (1997). *Machiavellian Intelligence II: Extensions and Evaluations*. Cambridge: Cambridge UP.

Call, J. and M. Tomasello (2008). 'Does the chimpanzee have a theory of mind? 30 years later.' *Trends in Cognitive Sciences*, 12(5): 187–192.

Cargile, J. (1970). 'A note on 'iterated knowings''. *Analysis* 30: 151-155.

Carney, J., R. Wlodarski, and R. Dunbar. (2014). 'Inference or enaction? The impact of genre on the narrative processing of other minds'. *PloS One* 9: (12): e114172.

Carruthers, P. (2004). 'Practical reasoning in a modular mind'. *Mind and Language* 19: 259-278.

Cefalu, P. (2013). 'The burdens of mind reading in Shakespeare's *Othello*: A cognitive and psychoanalytic approach to Iago's theory of mind.' *Shakespeare Quarterly* 64: (3): 265–294.

Chisholm, R. (1967). 'Intentionality.' In: P. Edwards (ed.). *The Encyclopedia of Philosophy*. Volume 5. London: Macmillan, pp. 201–204.

Cienki, Alan. (2007). 'Frames, Idealized Cognitive Models, and Domains.' In: D. Geeraerts and H. Cuyckens (eds.). *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford UP, pp. 170-178.

Clark, H.H. 1996. *Using Language*. Cambridge: Cambridge UP.

Clark, H.H. (2006). 'Social actions, social commitments.' In: N. J. Enfield and S. C. Levinson (eds.). *Roots of human sociality: Culture, cognition and interaction*. Oxford: Berg, pp. 39-69.

Clark, H.H. (2006a). 'Context and common ground.' In: K. Brown (ed.). *Encyclopedia of Language & Linguistics*. (Second Edition). Oxford: Elsevier, Volume 3, pp. 105-108.

Clark, H.H. and C.R. Marshall. (1981). 'Definite reference and mutual knowledge.' In: A.K. Joshi, B. Webber, and I. Sag (eds.). *Elements of Discourse Understanding*. Cambridge: Cambridge UP, pp. 10-63.

Clayton, N.S., J. M. Dally, and N. J. Emery. (2007). 'Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist'. *Trans R Soc Lond B Biol Sci*. 362(1480): 507–522.

Corballis, M.C. (2011). *The Recursive Mind. The Origins of Human Language, Thought, and Civilization*. Princeton: Princeton UP.

Coulson, S. (2001). *Semantic Leaps: Frame-shifting and Conceptual Blending in Meaning Construction*. New York: Cambridge UP.

Dancygier, B. (2008). 'The text and the story: levels of blending in fictional narratives'. In: T. Oakley and A. Hougaard. (eds.). (2008). *Mental spaces in discourse and interaction*. Amsterdam/Philadelphia: John Benjamins, pp. 51-78.

Dancygier, B. (2011). 'Modification and constructional blends in the use of proper names.' *Constructions and Frames* 3:(2): 208–235.

Dancygier, B. (2012). *The Language of Stories. A Cognitive Approach*. New York: Cambridge UP.

Dancygier, B. (2012a). 'Negation, stance verbs, and intersubjectivity.' In: Dancygier and Sweetser, 2012: 69-93.

Dancygier, B. and E. Sweetser. (2012). *Viewpoint in Language. A Multimodal Perspective*. Cambridge: Cambridge University Press.

David-Barrett, T. (2014). 'Language evolution bypasses the collective action paradox.' EHBEA Conference, April 9th, University of Bristol.

David-Barrett, T. and R.I.M. Dunbar. (2013). 'Processing Power Limits Social Group Size: Computational Evidence for the Cognitive Costs of Sociality.' *Proceedings of the Royal Society B* 280: 20131151.

De Jong, I.J.F., see Jong, I.J.F. de

De Villiers, J.G., see Villiers, J.G. de

De Waal, F.B., see Waal, F.B. de

Dennett, D. C. (1987). *The Intentional Stance.* Cambridge MA: MIT Press.

Dennett, D.C. (1971). 'Intentional Systems'. *Journal of Philosophy* 68: 87-106.

Dennett, D.C. (1983). 'Intentional systems in cognitive ethology: the 'Panglossian Paradigm' Defended.' *Behavioral and Brain Sciences* 6: 343–390.

Dennett, D.C. (1991). *Consciousness Explained*. Boston: Little, Brown and Co.

Dennett, D.C. (2000). Making Tools for Thinking. In D. Sperber (ed.). *Metarepresentations*. New York: Oxford UP, pp. 17-30.

Dennet, D.C. (2013). *Intuition Pumps and Other Tools for Thinking*. New York: W.W. Norton & Company.

Di Pellegrino, G., L. Fadiga, L. Fogassi, V. Gallese, G. Rizzolatti. (1992). 'Understanding motor events: a neurophysiological study.' *Experimental Brain Research* 91: (1): 176-180.

Diessel, H. and M. Tomasello. (2001). 'The acquisition of finite complement clauses in English: a corpus-based analysis.' *Cognitive Linguistics* 12: 97-141.

Djikic, M., K. Oatley, and M.C. Moldoveanu. (2013). 'Reading other minds: Effects of literature on empathy'. *Scientific Study of Literature* 3: (1): 28–47.

Du Bois, J.W. (1985). 'Competing Motivations.' In: J. Haiman (ed.). *Iconicity in Syntax*. Amsterdam: Benjamins, pp. 343-365.

Duijn, M.J. van. (2015). 'Lezen en gedachtenlezen. Verbetert literatuur ons inlevingsvermogen?' *Vooys* 33: (3):

Duijn, M.J. van, A. Rotkirch, T. David-Barrett, M. Jokela and R.I.M. Dunbar. (2014). 'Multiple-Order Intentionality Revisited.' EHBEA Conference, April 6th, University of Bristol.

Duijn, M.J. van, I. Sluiter, and A. Verhagen. (2015). 'When narrative takes over: The representation of embedded mindstates in Shakespeare's *Othello*'. *Language and Literature* 24: (2): 148–166.

Dunbar, R.I.M. (1992). 'Neocortex size as a constraint on group size in primates.' *Journal of Human Evolution* 22: 469-493.

Dunbar, R.I.M. (1998). 'The social brain hypothesis.' *Evolutionary Anthropology* 6:178-190.

Dunbar, R.I.M. (2003). 'The Social Brain: Mind, Language, and Society in Evolutionary Perspective'. *Annual Review of Anthropology* 32: 163-81.

Dunbar, R.I.M. (2005). 'Why are good writers so rare? An Evolutionary Perspective on Literature'. *J Cult Evol Psychol* 3: 7–21

Dunbar, R.I.M. (2008). 'Mind the Gap or Why Human Aren't Just Great Apes'. *Proceedings of the British Academy* 154: 403-23.

Dunbar, R.I.M. (2014). *Human Evolution*. London: Penguin Books.

Dunbar, R.I.M. (2015). 'Sexual segregation in human conversations.' *Behaviour* 152.

Dunbar, R., C. Gamble, and J. Gowlett (eds.). (2010). *Social brain, distributed mind*. Proceedings of the British Academy 158. Oxford UP.

Dunbar, R.I.M, R. Baron, A. Frangou (et al.) (2011). 'Social laughter is correlated with an elevated pain threshold.' *Proceedings of the Royal Society of London B: Biological Sciences* 20111373.

Eggers, D. (2000). *A Heartbreaking Work of Staggering Genius*. New York: Simon & Schuster.

Emery, N.J. and N.S. Clayton. (2004). 'The mentality of crows: convergent evolution of intelligence in corvids and apes.' *Science* 306: (5703): 1903-1907.

Evans, N. (2010). *Dying Words: Endangered languages and what they have to tell us*. Chichester: Wiley-Blackwell.

Evans, N. and S.C. Levinson. (2009). 'The myth of language universals: language diversity and its importance for cognitive science.' *Behavioral and Brain Sciences* 32: 429-492.

Fauconnier, G. (1985). *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge UP.

Fauconnier, G. (1997). *Mappings in Thought and Language*. Cambridge: Cambridge UP.

Fillmore, C.J. (1976). 'Frame semantics and the nature of language.' *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech* 280: 20–32.

Fillmore, C.J. (1981). 'Pragmatics and the description of discourse.' In: P. Cole (ed.). *Radical Pragmatics*. New York: Academic Press, pp. 143-166.

Fillmore, C.J. (1985). 'Frames and the semantics of understanding.' *Quaderni di Semantica* 6: 222-254.

Fitch, W.T. (2010). *The evolution of language*. New York: Cambridge University Press.

Fludernik, M. (1993). *The Fictions of Language and the Languages of Fiction: The Linguistic Representation of Speech and Consciousness*. London: Rouledge.

Fodor, J.A. (1983). *The Modularity of the Mind*. Cambridge MA: MIT Press.

Forbes, G. (1994). *Modern Logic. A Text in Elementary Symbolic Logic*. Oxford: Oxford UP.

Frith, C.D. and U. Frith. (2006). 'How we predict what other people are going to do.' *Brain Research* 1079: (1): 36–46.

Gallagher, S. and D.D. Hutto. (2008). 'Primary interaction and narrative practice.' In: J. Zlatev, T. Racine (et al.) (eds.). *The Shared Mind: Perspectives on Intersubjectivity.* Amsterdam and Philadelphia, PA: John Benjamins Publishing Company, pp. 17–38.

Gallagher, S. (2008). 'Inference or interaction: social cognition without precursors.' *Philosophical Explorations* 11: (3): 163–174.

Gallagher, S. (2009). 'Neural simulation and social cognition.' In: J.A. Pineda (ed.) *Mirror Neuron Systems: The Role of Mirroring Processes in Social Cognition*. Totowa NJ: Humana Press, pp. 355–371.

Gallagher, S. (2012). 'Empathy, Simulation, and Narrative.' *Science in Context* 25: 355-381.

Gallese, V. (2001). 'The "shared manifold" hypothesis: From mirror neurons to empathy.' *Journal of Consciousness Studies* 8: 33–50.

Gallese, V. and A. Goldman. (1998). 'Mirror neurons and the simulation theory of mind-reading.' *Trends in Cognitive Sciences* 12: 493–501.

Gamble, C. (2010). 'Technologies of Separation and the Evolution of Social Extension.' In: R. Dunbar, C. Gamble, and J. Gowlett (eds.). (2010). 'Social brain, distributed mind.' *Proceedings of the British Academy* 158.

Gamble, C., J. Gowlett, and R. Dunbar. (2014). *Thinking Big. How the Evolution of Social Life Shaped the Human Mind*. London: Thames and Hudson.

Genette, G. (1972). *Figures III*. Paris: Seuil.

Geurts, B. and P. Rubio-Fernández. (2015). 'Pragmatics and processing'. *Ratio* 12.

Gigerenzer, G., R. Hertwig, and T. Pachur (eds.). (2011). *Heuristics: The foundations of adaptive behavior*. New York: Oxford University Press.

Goldman, A. (2006). *Simulating minds: The philosophy, psychology and neuroscience of mindreading*. Oxford: Oxford University Press.

Gopnik, A. and H.M. Wellman. (2012). 'Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory'. *Psychological Bulletin* 138: (6): 1085–108.

Greenbaum, S. (1969). *Studies in English Adverbial Usage*. London: Longmans.

Grice, H.P. (1957). 'Meaning'. *The Philosophical Review* 66: (3): 377-388.

Grice, H.P. (1969). 'Utterer's meaning and intention. *The Philosophical Review* 78: (2): 147-177.

Hacker, P.M.S. (1986). *Insight and illusion. Themes in the Philosophy of Wittgenstein*. (Revised Edition). Oxford: Clarendon Press.

Hare, B., J. Call, and M. Tomasello. (2006). 'Chimpanzees deceive a human competitor by hiding.' *Cognition* 101: (3): 495-514.

Hall, E.T. (1966). *The Hidden Dimension*. Garden City, New York: Doubleday.

Helms, N.R. (2012). 'Conceiving ambiguity: Dynamic mindreading in Shakespeare's *Twelfth Night*.' *Philosophy and Literature* 36: (1): 122–135.

Henzi, S. P., L.F. de Sousa Pereira, D. Hawker-Bond, J. Stiller, R.I.M. Dunbar, and L. Barrett. (2007). 'Look who's talking: developmental trends in the size of conversational cliques'. *Evolution and Human Behavior* 28: (1): 66–74.

Herman, D. (2009). *Basic Elements of Narrative*. Chichester: Wiley-Blackwell.

Heyes, C.M. (1993). 'Imitation without perspective-taking.' *Behavioral and Brain Sciences* 16: 524-525.

Heyes, C. M. (2012). 'Simple minds: A qualified defence of associative learning'. *Philosophical Transactions of the Royal Society B* 367: 2695-2703.

Heyes, C.M. (2012a). 'What's social about social learning?' *Journal of Comparative Psychology* 126: 193-202.

Heyes, C. M. (2014). 'Submentalizing: I'm not really reading your mind'. *Perspectives on Psychological Science* 9: 131-143.

Heyes, C.M. and C.D. Frith. (2014). 'The cultural evolution of mind reading'. In: *Science* 344: 6190: 1243091.

Hühn, P. and R. Sommer. (2009). 'Narration in poetry and drama.' In: P. Hühn, J. Pier (et al.) (eds) *Handbook of Narratology*. Berlin: Walter de Gruyter, pp. 228–241.

Hühn, P., W. Schmid, and J. Schönert (eds.). (2009). *Point of View, Perspective, and Focalization: Modeling Mediation in Narrative*. Berlin: Walter de Gruyter.

Humphrey, G. (1924). 'The psychology of the gestalt'. *Journal of Educational Psychology*, 15: (7): 401–412.

Humphrey, N.K. (1976). 'The social function of intellect.' In: P.P.G. Bateson and R.A. Hinde (eds.). *Growing Points in Ethology*. Cambridge: Cambridge University Press, pp. 303–317.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Hutchins, E. (2006). 'The distributed cognition perspective on human interaction.' In: N. J. Enfield and S. C. Levinson (eds.). *Roots of human sociality: Culture, cognition and interaction*. Oxford: Berg, pp. 39-69.

Hutto, D.D. (2008). *Folk Psychological Narratives: The Socio-Cultural Basis of Understanding Reasons*. Cambridge MA: MIT Press.

Jeffries, L. and D. McIntyre. (2010). *Stylistics*. Cambridge: Cambridge UP.

Jolly, A. (1966). 'Lemur social behaviour and primate intelligence.' *Science* 153: 501–506.

Jong, I.J.F. de. (1997). 'Homer and narratology.' In: I. Morris and B. Powell (eds). *A New Companion to Homer*. Leiden: Brill, pp. 313–319.

Jong, I.J.F. de. (2009). 'Metalepsis in Greek Literature'. In: Jonas Grethlein, Antonios Rengakos (eds.). *Narratology and Interpretation: The Content of Narrative Form in Ancient Literature. Trends in Classics - Supplementary Volume*s 4. Berlin/New York: Walter de Gruyter, pp. 630-644.

Kahneman, D. and A. Tversky. (1982). 'The simulation heuristic'. In: D. Kahneman, P. Slovic, and A. Tversky (eds.). *Judgment Under Uncertainty*. Cambridge: Cambridge UP, pp. 201-208.

Keller, R. (1995). *Zeichentheorie. Zu einer Theorie semiotischen Wissens*. Tübingen/Basel: Franke.

Kidd, D.C. en E. Castano. (2013). 'Reading literary fiction improves theory of mind'. *Science* 342: 377–380.

Kilgarriff, A., V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý, V. Suchomel. (2014). 'The Sketch Engine: ten years on.' *Lexicography* 1: (1): 7-36.

Kinderman, P., R.I.M. Dunbar, and R.P. Bentall. (1998). 'Theory-of-mind deficits and causal attributions'. *British Journal of Psychology* 89: 191-204.

Knoblich, G. and N. Sebanz. (2006). 'The social nature of perception and action.' *Current Directions in Psychological Science* 15: (3): 99–104.

Koopman, E.M. and F. Hakemulder. (2015). 'Effects of Literature on Empathy and Self-Reflection: A Theoretical-Empirical Framework.' *Journal of Literary Theory* 9: (1): 79-111.

Korsten, F.W. (2005). *Lessen in Literatuur*. Amsterdam: Vantilt.

Kümmerli, R., M. Burton-Chellew, A. Ross-Gillespie, and S.A. West. (2010). 'Resistance to extreme strategies, rather than prosocial preferences, can explain human cooperation in public goods games.' *Proceedings of the National Academy of Sciences* 107: (22): 10125-10130.

Lakoff, G. (1977). 'Linguistic Gestalts.' *Chicago Linguistic Society* 13: 236-287.

Langacker, R.W. (1987). *Foundations of Cognitive Grammar,* ii. *Descriptive Application*. Stanford: Stanford UP.

Langacker, R.W. (1990). *Concept, Image, Symbol: The Cognitive Basis of Grammar.* Berlin: Mouton de Gruyter.

Langacker, R.W. (1990a). 'Subjectification.' *Cognitive Linguistics* 1: 1: 5-38.

Launay, J., E. Pearce, R. Wlodarski, M.J. van Duijn, J. Carney, and R.I.M. Dunbar. (2015). 'Higher-order mentalising and executive functioning'. *Psychology and Individual Differences* 86: 6–14.

Leech, G. and M. Short. (2007). *Style in Fiction. A Linguistic Introduction to English Fictional Prose*. Second Edition. Harlow: Pearson Education Limited.

Leslie, A., O. Friedman, and T.P. German. (2004). 'Core mechanisms in "theory of mind".' *Trends in Cognitive Sciences* 8: (12): 528–533.

Levinson, S.C. (2006). 'On the human "interaction engine".' In: N. J. Enfield and S. C. Levinson (eds.). *Roots of human sociality: Culture, cognition and interaction*. Oxford: Berg, pp. 39-69.

Levinson, S.C., Evans, N. (2010), 'Time for a sea-change in linguistics: Response to comments on 'The Myth of Language Universals'. *Lingua* 120: 2733-2758.

Lewis, D.K. (1969). *Convention: A Philosophical Study*. Cambridge, MA: Harvard UP.

Lewis, P. A., R. Rezaie, R. Brown, N. Roberts, and R.I.M. Dunbar. (2011). 'Ventromedial prefrontal volume predicts understanding of others and social network size.' *NeuroImage*, 57: (4): 1624–1629.

Liddle, B. and D. Nettle. (2006). 'Higher-order theory of mind and social competence in school-age children.' *Journal of Cultural and Evolutionary Psychology* 4: (3–4): 231–246.

Lohmann, H. and M. Tomasello. (2003). 'The role of language in the development of false belief understanding: A training study.' *Child Development* 74: (4): 1130–1144.

Lyons, M., T. Caldwell, and S. Shultz. (2010). 'Mind-reading and manipulation - Is Machiavellianism related to theory of mind?' *Journal of Evolutionary Psychology* 8:(3): 261–274.

Maguire, L. (2014). *Othello. Language and Writing*. London: Bloomsbury Arden Shakespeare.

Marr, D. (1982). *Vision: A Computational Approach*. San Francisco: Freeman & Co.

Mars, R.B., J. Sallet, F.X. Neubert, and M.F.S. Rushworth. (2013). 'Connectivity profiles reveal the relationship between brain areas for social cognition in human and monkey temporoparietal cortex.' *PNAS* 110: (26): 10806–10811.

McConachie, B. and F.E. Hart. (eds.). (2006). *Performance and Cognition. Theatre Studies and the Cognitive Turn*. London and New York: Routledge.

McGregor, William B. (1997). *Semiotic Grammar*. Oxford/New York: Oxford UP.

Mesoudi, A. (2011). *Cultural Evolution: How Darwinian theory can explain human culture and synthesize the social sciences*. Chicago, IL: University of Chicago Press.

Mey, J.L. (1999). *When voices clash*. Berlin/New York: Mouton De Gruyter.

Milligan, K., J.A. Astington, and L.A. Dack. (2007). 'Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding.' *Child Development* 78: (2): 622–646.

Moles, A. (1963). 'Animal language and information theory.' In: R. Busnel (ed.). *Acoustic Behaviour of Animals.* Amsterdam, Elsevier Publishing, pp. 112-131.

Moll, H. and M. Tomasello. 'Cooperation and human cognition: the Vygotskian intelligence hypothesis.' *Phil. Trans. R. Soc. B* 362: 639-648.

Murakami, H. (2000). *Norwegian Wood*. New York: Vintage Books/Random House.

Murakami, H. (2014). *Colorless Tsukuru Tazaki and His Years of Pilgrimage*. New York: Vintage Books/Random House.

Nettle, D., and B. Liddle. (2008). 'Agreeableness is related to social-cognitive, but not social- perceptual, theory of mind.' *European Journal of Personality* 22: 323–335.

Nussbaum, M. (1997). *Cultivating Humanity. A Classical Defense of Reform in Liberal Education*. Cambridge, MA: Harvard UP.

O'Grady, C., C. Kliesch, K. Smith, and T.C. Scott-Phillips. (2015). 'The ease and extent of recursive mindreading, across implicit and explicit tasks.' *Evolution and Human Behavior* 10: 1016.

Oatley, K. (2011). *Such stuff as dreams: The psychology of fiction.* Oxford: Wiley Blackwell.

Oesch, N.T. (2015). *The Adaptive Significance of Human Language. Function, Form, and Social Cognition*. PhD Thesis, Oxford University.

'Oscar Pistorius shoots girlfriend: report.' SAPA [South African Press Association], 14 February 2013. Accessed online through http://academic.lexisnexis.nl.

Owings, DH. and E.S. Morton. (1998). *Animal Vocal Communication. A New Approach*. Cambridge: Cambridge UP.

Palmer, A. (2004). *Fictional Minds*. Lincoln: U of Nebraska P.

Palmer, A. (2010). 'Storyworlds and Groups'. In: Zunshine, 2010: 167-192.

Papp, S. (2006). 'A relevance-theoretic account of the development and deficits of theory of mind in normally developing children and individuals with autism.' *Theory & Psychology* 16: (2): 141–161.

Parr, L.A. and F.B. de Waal. 'Visual kin recognition in chimpanzees.' (1999). *Nature* 399: 647-648.

Pearce, E., J. Launay, M.J. van Duijn, T. David-Barrett, A. Rotkirch, and R.I.M. Dunbar. (forthcoming). 'Singing together or apart: The effect of competitive and cooperative singing on social bonding within and between sub-groups of a University Fraternity.' *Psychology of Music*.

Perreault, C. and S. Mathew. (2012). 'Dating the origin of language using phonemic diversity.' *PLoS ONE* 7: (4): e35289.

Pickering, M.J. and S. Garrod. (2004). 'Towards a mechanistic psychology of dialogue.' *Behavioural and Brain Sciences* 27: 169-226.

Pineda, J.A. (2009). *Mirror Neuron Systems: The Role of Mirroring Processes in Social Cognition*. Humana Press.

Pinkster, H. 1990. *Latin Syntax and Semantics.* London/New York: Routledge.

Powell, J. L., P. Lewis, R.I.M. Dunbar, M. García-Fiñana, and N. Roberts. (2010). 'Orbital prefrontal cortex volume correlates with social cognitive competence.' *Neuropsychologia*, 48: 12: 3554–3562

Powell, J. L., P. Lewis, N. Roberts, M. García-Fiñana, and R.I.M. Dunbar. (2012). 'Orbital prefrontal cortex volume predicts social network size: An imaging study of individual differences in humans.' *Proceedings of the Royal Society B: Biological Sciences* 279: (1736): 2157–2162.

Powell, J. L., G.J. Kemp, R.I.M. Dunbar, N. Roberts, V. Sluming, and M. García-Fiñana. (2014). 'Different association between intentionality competence and prefrontal volume in left- and right-handers.' *Cortex* 54: 63–76.

Premack, D.G. and G. Woodruff. (1978). 'Does the Chimpanzee Have a Theory of Mind?' *Behavioral and Brain Sciences* 1: 515-526.

Ridley, M.R. (1965). *Othello.* London: Arden Shakespeare Paperbacks/Methuen & Co.

Rosati, A. and B. Hare. (2010). 'Social cognition: From behavior-reading to mind-reading.' In: G. Koob, M. Le Moal, and R. Thompson (eds.). *Encyclopedia of Behavioral Neuroscience*. Erlbaum 3, pp. 263-268.

Rushworth, M.F.S., R.B. Mars, and J. Sallet. (2013). 'Are there specialized circuits for social cognition and are they unique to humans?' *Current Opinion in Neurobiology* 23: 436-442.

'S.Africa's Pistorius arrested after killing girlfriend: report.' Agence France Presse, 14 February 2013. Accessed online through http://academic.lexisnexis.nl.

Sallet, J., R.B. Mars, M.P. Noonan, J.L. Andersson, J.X. O'reilly, S. Jbabdi (et al.). (2011). 'Social network size affects neural circuits in macaques.' *Science* 334: (6056): 697-700.

Saxe, R. (2006). 'Uniquely human social cognition'. *Current Opinion in Neurobiology* 16: 235-239.

Schaafsma, S.M., D.W. Pfaff, R.P. Spunt, and R. Adolphs. (2015). 'Deconstructing and reconstructing theory of mind'. *Trends in Cognitive Sciences* 19: 2: 65-72.

Schneider, J. (2003). *Einführung in die Roman-Analyse*. Darmstadt: Germanistik Perfect Paperback.

Scott-Phillips, T.C. (2015). *Speaking Our Minds. Why Human Communication is Different and How Language Evolved to Make it Special*. New York: Palgrave Macmillan.

Seal, M. (2013). 'Inside the Oscar Pistorius Murder Case: What Paths Took the Couple to That Final Night?' *Vanity Fair,* June 1. http://www.vanityfair.com/culture/2013/06/oscar-pistorius-murder (accessed 4 March 2014).

Searle, J.R. (1969). *Speech acts. An essay in the philosophy of language.* Cambridge: Cambridge UP.

Searle, J.R. (1983). *Intentionality: An Essay in the Philosophy of Mind.* Cambridge: Cambridge University Press.

Searle, J.R. (1991). *The Rediscovery of the Mind*. Cambridge, Mass.; London: MIT Press.

Searle, J.R. (1995). *The Construction of Social Reality*. New York: Free Press.

Seyfarth, R.M., D.L. Cheney, and P. Marler. (1980). 'Vervet monkey alarm calls: semantic communication in a free-ranging primate.' *Animal Behaviour* 28:(4): 1070-1094.

Shultz, S. and R.I.M. Dunbar. (2007). 'The evolution of the social brain: Anthropoid primates contrast with other vertebrates'. *Proceedings of the Royal Society, London* 274B: 2429-2436.

Shultz, S., and Dunbar, R. I.M. (2014). 'The social brain hypothesis: An evolutionary perspective on the neurobiology of social behavior.' In: R. I. M. Dunbar, C. Gamble, and J. A. J. Gowlett (eds.), *Lucy to Language: the Benchmark Papers*. Oxford: Oxford University Press, pp. 53–70.

Shultz, S., C. Opie, and Q.D. Atkinson. (2011). 'Stepwise evolution of stable sociality in primates.' *Nature* 479: 219-222.

Sluiter, I., B. Corthals, M.J. van Duijn, and M.J.O. Verheij. (2013). In het hoofd van Medea. Gedachtenlezen bij een moordende moeder. *Lampas* 46: (1): 3-20.

Sorem, E. (2010). 'Searle, Materialism, and the Mind-Body Problem'. *Perspectives: International Journal of Philosophy* 3.

Sperber, D. (1994). 'Understanding Verbal Understanding'. In: J. Khalfa (ed.). *What is Intelligence?* Cambridge: Cambridge UP, pp. 179-98.

Sperber, D. (2000). 'Metarepresentations in an evolutionary perspective'. In: D. Sperber (ed.). *Metarepresentations: an Interdisciplinary Perspective*. Oxford: Oxford UP, pp. 117-137.

Sperber, D. and D. Wilson. (1986). *Relevance*. Oxford: Blackwell.

Sperber, D. and D. Wilson. (1995). *Relevance: Communication and Cognition*. (Second Edition). Oxford: Basil Blackwell.

Sperber, D. and D. Wilson. (2002). 'Pragmatics, modularity, and mindreading.' *Mind and Lanuage* 17: 3-23.

Stalnaker, R.C. (1978). 'Assertion.' In: P. Cole (ed.). *Syntax and Semantics 9: Pragmatics*. New York: Academic Press, pp. 315-332.

Stiller, J. and R.I.M. Dunbar. (2007). 'Perspective-Taking and Social Network Size in Humans'. *Social Networks* 29: 93-104.

Stiller, J., D. Nettle, and R. Dunbar. (2003). 'The small world of Shakespeare's plays.' *Human Nature* 14(4): 397–408.

Stolk, A. (2014). *On the generation of shared symbols.* PhD Thesis, Radboud Universiteit. Nijmegen: Donders Series 166.

Stolk, A., L. Verhagen, and I. Toni. (2016). 'Conceptual Alignment: How Brains Achieve Mutual Understanding. *TICS* 1521.

Sweeters, E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge UP.

Sweetser, E. (2006). Negative space: levels of negation and kinds of spaces. In: Stéphanie Bonnefille and Sébastian Salbayre (eds.). *Proceedings of the conference 'Negation: Form, figure of speech, conceptualization'*. Tours: Publications universitaires François Rabelais.

Sweetser, E. (2012). 'Introduction: viewpoint and perspective in language and gesture, from the Ground down.' In: Dancygier and Sweetser, 2012: 1-22.

Tarr, B., J. Launay, and R.I.M. Dunbar. (2014). 'Music and social bonding: "self- other" merging and neurohormonal mechanisms.' *Frontiers in Psychology* 5: 1–10.

Ter Meulen, Alice G.B. (2003). 'Speaking One's Mind.' In: Patrick Blackburn, Chiara Ghidini, Roy M. Turner, and Fausto Giunchiglia (eds.). *Modelling and Using Context* (4th International and Interdisciplinary Conference CONTEXT 2003, Stanford, CA, USA, June 23-25, 2003 Proceedings), Berlin & Heidelberg: Springer, pp. 507-516.

'Testosterone reported at home of Pistorius.' *International Herald Tribune*, 21 February 2013. Accessed online through http://academic.lexisnexis.nl.

'The One Where Everybody Finds Out'. (1999). In: D. Crane and M. Kauffman (prod.). *Friends*. Series 5, episode 14. Warner Bros Entertainment.

Tinbergen, N. (1963). 'On Aims and Methods of Ethology'. *Zeitschrift für Tierpsychologie* 20: 410–433.

Tomasello, M. (1999). *The cultural origins of human cognition.* Cambridge, MA: Harvard UP.

Tomasello, M. (2008). *Origins of Human Communication*. Boston: MIT Press.

Tomasello, M. (2014). *A Natural History of Human Thinking*. Cambridge/London: Harvard UP.

Tomasello, M. and H. Rakoczy. (2003). 'What makes human cognition unique? From individual to shared to collective intentionality'. *Mind and Language* 18: 121-147.

Toolan, M. 'Speech and Thought Representation.' In K. Brown (ed.). *The Encyclopaedia of Language and Linguistic*s (volume 11). Oxford: Elsevier, pp. 698-710.

Turner, M. and G. Fauconnier. (1995). 'Conceptual Integration and Formal Expression'. *Metaphor and Symbolic Activity* 10: (3): 183-203.

Uzefovsky, F., I. Shalev, S. Israel, S. Edelman, Y. Raz, D. Mankuta, A. Knafo-Noam, R.P. Ebstein. (2015). 'Oxytocin receptor and vasopressin receptor 1a genes are respectively associated with emotional and cognitive empathy.' *Hormones and Behavior* 67: 60-66.

Van Duijn, M.J., see Duijn, M.J. van

Vandelanotte, L. (2002). 'Prenominal adjectives in English: Structures and ordering.' *Folia Linguistica* XXXVI: (3-4): 219-259.

Vandelanotte, L. (2007). 'Mister so-called X. Discourse functions and subjectification of so-called.' In: C. Butler, R. Hidalgo, and J. Downing (eds.). *Functional perspectives on grammar and discourse: In honour of Angela Downing*, Amsterdam: Benjamins, pp. 359-394.

Vandelanotte, L. (2009). *Speech and Thought Representation in English*. Berlin and New York: Mouton de Gruyter.

Vandelanotte, L. (2012). '"Wait till you got started". How to submerge another's discourse in your own.' In: Dancygier and Sweetser, 2012: 198-218.

Verhagen, A. (2005). *Constructions of Intersubjectivity. Discourse, Syntax, and Cognition*. Oxford: Oxford University Press.

Verhagen, A. (2010). 'What do you think is the proper place of recursion? Conceptual and empirical issues.' In: H. van der Hulst (ed.). *Recursion and Human Language*. Berlin/New York: De Guyter Mouton, pp. 93-110.

Verhagen, A. (2015). 'Grammar and Cooperative Communication'. In: E. Dąbrowska en D. Divjak (eds.). *Handbook of Cognitive Linguistics* (Handbooks of Linguistics and Communication Science 39). Berlin & Boston, pp. 232-251.

Vermeule, B. (2010). *Why Do We Care about Literary Characters?* Baltimore: John Hopkins UP.

De Villiers, J.G. and P.A. De Villiers. (2003). 'Language for thought. Coming to understand false beliefs.' In: D. Gentner and S. Goldin-Meadow (eds.). *Language*

in Mind. Advances in the Study of Language and Thought. Cambridge MA: MIT Press, pp. 335–384.

Voßen, R. (1997). *Die Khoe-Sprachen: Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas*. (Volume 12). Köln: Rüdiger Köppe Verlag.

Vygotsky, L.S. (1978). *Mind in society: the development of higher psychological processes.* Cambridge, MA: Harvard UP.

Waal, F.B. de. (2005). *De aap in ons.* Amsterdam: Uitgeverij Contact.

Waal, F.B. de. (2013). *The Age of Empathy: Nature's Lessons for a Kinder Society*. New York: Harmony Books.

Warneken, F. and M. Tomasello. (2006). 'Altruistic helping in human infants and young chimpanzees.' *Science* 31: 1301–1303.

Whalen, D.H., L. Zunshine, and M. Holquist. (2012). 'Theory of Mind and embedding of perspective: A psychological test of a literary 'sweet spot''. *Scientific Study of Literature* 2: (2): 301–315.

Whiten, A. (2011). 'The scope of culture in chimpanzees, humans and ancestral apes' *Philosophical Transactions of the Royal Society B, Biological Sciences* 366: (1567): 997-1007.

Whiten, A. (2013). 'Culture and the evolution of interconnected minds.' In: Baron-Cohen et al., 2013, pp. 431-447.

Whiten, A., J. Goodall, W.C. McGrew, T. Nishida, V. Reynolds, Y. Sugiyama (et al.). (1999). 'Cultures in chimpanzees.' *Nature* 399: (6737): 682-685.

Wierzbicka, A. (2006). *English: Meaning and Culture*. Oxford: Oxford University Press.

Wilkins, D. (1986). 'Particle/Clitics for Criticism and Complaint in Mparntwe Arrernte (Aranda).' *Journal of Pragmatics* 10: 575-596.

Wilkinson, G.S. (1984). 'Reciprocal food sharing in the vampire bat.' *Nature* 308: (5955): 181-184.

Wilson, E.O. (1998). *Consilience: The Unity of Knowledge.* New York: Knopf.

'Witness heard 'non-stop shouting' before Pistorius shooting.' Reuters News, 20 February 2013.

Wittgenstein, L. (2006 [1953]). *Philosophische Untersuchungen*. Suhrkamp, Frankfurt am Main.

Woolf, V. (1925). *Mrs Dalloway*. London: Hogarth Press.

Woolf, V. (1966 [1924]). 'Mr. Bennett and Mrs. Brown'. In: *Collected essays*. [Vol. 1]. London: Chatto & Windus, 319–337.

Yamamoto, S., T. Humle, and M. Tanaka. (2012). 'Chimpanzees' flexible targeted helping based on an understanding of conspecifics' goals'. *PNAS* 109: (9): 3588–3592.

Yoshida, W., B. Seymour, K.J. Friston, and R.J. Dolan. (2010). 'Neural mechanisms of belief inference during cooperative games'. *Journal of Neuroscience* 30: (32): 10744–51.

Yoshida, W., R.J. Dolan, K.J. Friston. (2008). 'Game theory of mind.' *PLoS Comput Biol* 4: e1000254.

Zunshine L. (2011). 'What to expect when you pick up a graphic novel'. *SubStance* 40: 114–134.

Zunshine, L. (2006). *Why We Read Fiction. Theory of Mind and the novel.* Columbus: The Ohio State University Press.

Zunshine, L. (2010). *Introduction to Cognitive Cultural Studies*. Baltimore: John Hopkins UP.

Zunshine, L. (2012). 'Sociocognitive Complexity'. *NOVEL A Forum on Fiction* 45: (1): 13-18.

Samenvatting

De luie gedachtenlezer

Buren en naaste verwanten

Voor de deur van het huis waar ik sinds kort woon, stond ik recent een paar planken in te korten met een decoupeerzaag. Stofmasker voor, oorbeschermers op, werkhandschoenen, grote beschermbril, zaagsel in mijn haar. De buurman die ik al enkele keren met zijn hondje had zien lopen kwam zijn deur uit. Hij is een opvallende verschijning met lang, donkergeverfd haar en modieuze kleding. Met een gestileerde beweging stak hij twee duimen omhoog en zei: "Ziet er goed uit, buurman!"

Pas dagen later kwam ik erachter dat dit niet zomaar een (al dan niet ironisch bedoeld) complimentje van een buurtgenoot was. Iemand wees mij er namelijk op dat de buurman bekend is van televisie als helft van het illustere duo dat de "modepolitie" vormt. In het gelijknamige programma gaan de "agenten" van deze politie blijkbaar de straat op om voorbijgangers te betrappen op schrijnende modemissers, of ze juist te complimenteren met een geslaagd voorkomen. Nu *begreep* ik opeens dat de buurman waarschijnlijk *dacht* dat ik *wist* dat hij bekend was van de modepolitie, en dat ik daarom *zou verwachten* dat hij iets van mijn opmerkelijke klusoutfit *zou vinden*. Wat de buurman niet kon weten is dat ik een notoir slechte kenner ben van beroemde televisiegezichten.

Deze situatie heeft van alles met het onderwerp van mijn proefschrift te maken. Laten we ervan uitgaan dat de buurman inderdaad een grap maakte gebaseerd op zijn televisierol: hij dacht dat ik zou weten dat hij van de modepolitie was, en voerde daarom zijn voor veel kijkers direct herkenbare act uit met de duimen omhoog op het moment dat ik in vol doe-het-zelfornaat op straat stond. Om deze grap tot zijn volle recht te laten komen, is een proces nodig dat ik in mijn proefschrift *gedachtenlezen* noem. Immers, ik moet me verplaatsen in het perspectief van de buurman en begrijpen dat hij *denkt* dat ik

weet dat hij van de modepolitie is, en dat hij een act opvoert die in de context van dit programma *betekent* dat ik voldoe aan de momenteel geldende modestandaarden. Daarbij beseffend dat ik met mijn oorbeschermers, stofmasker en plastic laboratoriumbril op waarschijnlijk aan alles behalve deze standaarden voldoe, kan ik *vaststellen* dat de woorden van de buurman niet letterlijk moeten worden opgevat…

Dat we tot op zekere hoogte kunnen vaststellen wat er in andermans hoofd omgaat lijkt misschien vanzelfsprekend, maar over het geheel van de levende natuur genomen is ons vermogen tot gedachtenlezen absoluut spectaculair. Niet-menselijke dieren zijn hier vele malen slechter in: als ze het al kunnen, dan in veel minder diepte en detail. Het ligt voor de hand dat het gegeven dat andere dieren minder goed kunnen gedachtenlezen, er iets mee te maken heeft dat ze geen taal hebben. Maar die vaststelling op zichzelf is niet voldoende, ze verschuift het probleem alleen maar. Het menselijk taalvermogen is niet te begrijpen zonder inzicht in het geheel van mechanismen dat mensen gebruiken om zich in anderen te verplaatsen, en uit te vinden wat ze denken, bedoelen, hopen, van plan zijn, enzovoorts. Taaluitingen zijn immers bijna nooit letterlijke weergaven van gedachtentoestanden. Ze vormen een krachtig hulpmiddel als je probeert je gedachten met een ander af te stemmen, maar moeten door die ander altijd worden geïnterpreteerd en aangevuld met achtergrondkennis, wat een zekere mate van gedachtenlezen vraagt. Taalgebruikers moeten zich voortdurend (meer of minder bewust) realiseren wat hun gesprekspartner *eigenlijk* bedoelt met een bepaalde uitspraak en in het licht van welke gedeelde achtergrondkennis deze precies moet worden begrepen. Iemand die dit—vaak bij wijze van flauwe grap—even achterwege laat, zegt bijvoorbeeld "ja, dat mag" op de vraag of een tafelgenoot de peper zou mogen, in plaats van het molentje daadwerkelijk aan te reiken.

Iedereen kan talloze van zulke voorbeelden uit de dagelijkse omgang verzinnen en als je erop let, dan geldt voor nagenoeg alle talige communicatie dat de taaluitingen *op zichzelf* ("in isolatie" bezien) geen doorslaggevende rol spelen, maar een sturende werking hebben in een veel ruimer proces van het uitonderhandelen van betekenissen en het op elkaar afstemmen van gedachtentoestanden. Bij dit proces speelt de capaciteit tot gedachtenlezen een

belangrijke rol: gesprekspartners moeten zich tot op zekere hoogte in elkaar kunnen verplaatsen om te zien wat de ander bedoelt met een bepaalde uiting. Behalve taal dragen vaak ook gezichtsuitdrukkingen, lichaamshouding, blikrichting, enzovoorts, bij aan het afstemmingsproces. Met andere woorden, taaluitingen op zichzelf zijn maar een deel van het verhaal van communicatie, en moeten worden aangevuld met andere elementen die het tot stand brengen van wederzijds begrip mogelijk maken—denk maar aan het spelletje "Hints": als je geen taal mag gebruiken is het weliswaar lastiger om over te brengen wat je bedoelt, maar het lukt alsnog vrij aardig door gezichtsuitdrukkingen, lichaamshouding, enzovoorts, optimaal te benutten.

De gangbare opvatting in de wetenschappelijke literatuur over dit onderwerp is dat menselijke interactie van beide partijen vraagt om *recursief* te denken, ofwel *meervoudige metarepresentatie* toe te passen. Mensen moeten dus, volgens deze opvatting, voortdurend meerdere geestesgesteldheden ("intentional states") in elkaar kunnen inbedden zonder de draad kwijt te raken. Bij elke communicatieve interactie die A en B hebben moet A in staat zijn te *begrijpen* dat B *bedoelt* dat A *zal denken* dat X het geval is—met andere woorden, A moet niet alleen "decoderen" wat B zegt, maar ook voortdurend het perspectief van B proberen te nemen (c.q. zijn gedachten proberen te lezen) om te begrijpen wat B werkelijk met zijn boodschap wil zeggen. Onderzoekers als Dan Sperber (2000) en Thom Scott-Phillips (2015) gaan nog een stap verder en stellen dat zelfs dit niet genoeg is: volgens hen moeten taalgebruikers aan de zijde van de zender en ontvanger respectievelijk vier en vijf stappen van inbedding maken om normaal te kunnen communiceren. Komt er net als in de situatie met mijn buurman bovendien nog ironie bij kijken, dan komen er nog extra lagen bovenop en zitten sprekers en ontvangers zo op zes of zeven stappen:

De spreker *wil*
dat de luisteraar *begrijpt*
dat de spreker *bedoelt*
dat de luisteraar *inziet*
dat de spreker niet letterlijk X *voor ogen heeft*
maar eigenlijk Y *bedoelt*

(vgl. Hoofdstuk 5 van dit proefschrift en Scott-Phillips 2015; er bestaat discussie over de vraag hoe de inbeddingen precies moeten worden geconstrueerd in geval van ironie)

Hier raakt dit theoretische "model" van menselijke interactie een belangrijk punt uit een ander vakgebied. Onderzoekers uit de experimentele en evolutionaire psychologie zijn al sinds enkele decennia geïnteresseerd in de vraag hoeveel van deze lagen ("orders of intentionalilty") mensen en andere dieren aankunnen. De basisgedachte is dat veruit de meeste dieren een cognitief systeem van de eerste orde belichamen: ze hebben allerlei gedachten zoals "eetbaar", "giftig", "die kant op", "vijand", enzovoorts (dit wil overigens niet zeggen dat ze zich *bewust* zijn van deze gedachten zoals wij mensen dat kunnen zijn). Maar wat ze niet hebben zijn gedachten over gedachten: "hij *wil* mijn eten stelen", "mijn kind *denkt* dat ik wegga", enzovoorts—dit is voorbehouden aan cognitieve systemen van de tweede orde. De algemene opvatting is dat geen enkel ander levend wezen dan de mens drie of meer ordes aankan, mogelijk met als enige uitzondering onze naaste verwanten in de natuur, chimpansees en bonobo's. Maar zelfs voor deze mensapen geldt dat ze slechts bij uitzondering slagen voor testjes waarin het nodig is te redeneren met drie ordes van ingebedde geestesgesteldheden.

Dit is vooral interessant aangezien chimpansees en bonobo's een heel belangrijke rol spelen in het onderzoek naar de evolutie van de mens. Tussen vijf en zeven miljoen jaar geleden leefde in Afrika rondom de evenaar een primaat die de voorouder was van zowel de hedendaagse chimpansees en bonobo's als van de moderne mens, onze zogenaamde "last common ancestor" of LCA. Met name de chimpansee wordt vaak gezien als een goed model voor hoe deze voorouder er uitzag en leefde. Het is dus redelijk om aan te nemen dat de LCA qua gedachtenlezen op hetzelfde niveau zat als de hedendaagse chimpansee—op twee, in sommige gevallen drie ordes, uitgaande van de bovenstaande manier van tellen. Van onszelf, de hedendaagse mens, weten we dat we vanaf een bepaalde leeftijd slagen voor testjes waarin met vier, vijf of soms zelfs zes van zulke ordes van ingebedde geestesgesteldheden moet worden geredeneerd. Dit suggereert dat mens(achtig)en in de loop van hun evolutionaire pad dus een steeds geavanceerder vermogen tot gedachtenlezen

moeten hebben ontwikkeld. Gegeven de juiste context en presentatie (daarover zo meteen meer) kunnen wij *begrijpen* dat ik me *realiseer* dat de buurman *dacht* dat ik *verwachtte* dat hij *vond* dat mijn outfit er niet uitzag. Op basis van experimenten (waarover hieronder ook meer) wordt in het algemeen aangenomen dat dit de limiet is waarop ons evolutionaire pad is gestrand: waar sommige dieren op eerste orde zitten, andere op tweede of misschien derde, zitten mensen op vijfde of misschien zesde orde.

De suggestie dat mensen een hogere-orde cognitief systeem belichamen en andere mensapen niet, zou kunnen stroken met een direct zichtbaar verschil: onze herseninhoud is ongeveer vier keer zo groot. Robin Dunbar (e.g. 2014), die een invloedrijke positie inneemt in dit debat, verklaart de groei van ons brein in de afgelopen paar miljoen jaar als een reactie op de steeds grotere sociale complexiteit waar onze voorouders mee te maken kregen. Deze verklaring staat bekend als de *social brain hypothesis*. Volgens hem zijn we in steeds grotere en fijnzinniger georganiseerde groepen gaan leven: waar chimpansees (vermoedelijk net als onze gemeenschappelijke voorouders) in hun natuurlijke omgeving leven in verbanden van 40-60 individuen, lijkt archeologische evidentie te suggereren dat dit aantal bij mensen gaandeweg gegroeid is naar ongeveer 150—de groepsgrootte die we nu nog vinden bij hedendaagse jager-verzamelaarvolkeren. Dit lijkt misschien een geringe stijging, maar bedenk dat primatengroepen een netwerkstructuur vertonen. Anders dan bijvoorbeeld in een kudde schapen of een vlucht vogels is het niet zo dat individuen een aantal willekeurige (en regelmatig wisselende) buren hebben waarmee ze rekening moeten houden, maar in een primatengroep heeft elk individu een persoonlijke relatie met ieder ander. Leden moeten bovendien iets weten over de onderlinge relaties van anderen om succesvol te kunnen opereren binnen het netwerk: wie is familie van wie? Wie zijn bevriend en wie staan op gespannen voet? Komt er in een groep van 50 één nieuw lid bij, dan moet ieder individu niet alleen een relatie ontwikkelen met dit ene nieuwe lid, maar ook in de gaten houden wat de relatie van de andere 50 leden is tot de nieuwkomer. Daarom is het goed verdedigbaar dat de hoeveelheid cognitief vermogen die nodig is om in een primatennetwerk te kunnen functioneren exponentieel stijgt met het aantal leden. Op basis hiervan doet Dunbar de suggestie dat de afhankelijkheden, globaal samengevat, als volgt zijn: er bestaat

druk op populaties van onze voorouders om in grotere groepen te gaan leven (bijvoorbeeld omdat dit veiliger is). Vervolgens oefent dit druk uit op individuen om zich beter in deze groepen te kunnen handhaven. Wie een iets krachtiger brein heeft—waarmee hij iets beter kan redeneren over wat anderen in de groep denken, willen, bedoelen, enzovoorts, inclusief wat anderen denken dat weer anderen denken—kan zich beter handhaven in een grotere groep en heeft meer kans om zich succesvol voort te planten. Zo ontstaat er een stijgende lijn in het vermogen tot gedachtenlezen, fysiek zichtbaar in de groei van het brein. Maar een groter brein op zichzelf verklaart niet hoe wij ingewikkelde gedachtenleestaken oplossen, zoals die in het voorbeeld van mijzelf en de buurman. Wat doen we precies als we met zo'n situatie te maken krijgen; of, in het groot gezegd: hoe werkt het sociale brein? Ik licht eerst de methoden, onderzoeksmaterialen en analyses uit mijn proefschrift toe. Vervolgens leg ik uit wat deze analyses te maken hebben met gedachtenlezen, het sociale brein en onze evolutie, en tot slot kom ik op terug op het voorbeeld.

Taal en verhalen ondersteunen gedachtenlezen

Verschillende onderzoekers (met name Sperber en Scott-Phillips) hebben dus beweerd dat het voor normale interactie nodig is om met vijf ordes van ingebedde gedachtentoestanden te kunnen werken. Ze verklaren op deze manier ook waarom menselijke communicatie uniek is: volgens hen was het vermogen tot vijfde-orde gedachtenlezen de basis waarop taal kon ontstaan. Ik betoog in mijn proefschrift dat deze opvatting verschillende problemen oplevert. Ten eerste lijkt het onwaarschijnlijk dat normale interactie al uitgaat van wat volgens sommigen de limiet is van ons vermogen tot gedachtenlezen: wat gebeurt er dan als er nog meer ordes bijkomen omdat er bijvoorbeeld ironie in het spel is, of omdat een gesprek gaat over wat persoon C denkt dat D van plan is E te doen hopen? Ten tweede blijkt uit de genoemde testjes dat hogere-orde gedachtenlezen een cognitief inspannende activiteit is. Hoe valt dit te verenigen met het feit dat we de hele dag door met alle gemak taal gebruiken? Ten derde zijn er aanwijzingen dat kinderen eerst bepaalde stappen in hun taalontwikkeling doormaken voordat ze in staat zijn tot hogere-orde gedachtenlezen. Dit suggereert dat ook het evolutionaire scenario het

overwegen waard is waarin taal aan hogere-orde gedachtenlezen voorafgaat (in plaats van dat taal noodzakelijkerwijs erop volgt). Ten vierde lijkt analyse van verschillende genres van taalgebruik in contexten waar meerdere perspectieven een rol spelen erop te wijzen dat taal, en in het bijzonder narratief taalgebruik, het moeten redeneren met meerdere ingebedde perspectieven juist "van ons overneemt". Wat ik hier precies mee bedoel, zal duidelijk worden aan het eind van deze paragraaf.

In Hoofdstuk 2 begin ik met het bespreken van een toneelstuk waarin het publiek in korte tijd meerdere ingebedde perspectieven voor de kiezen krijgt: Shakespeare's *Othello*. Al aan het eind van de tweede akte moet een toeschouwer of lezer van het stuk inzien dat

> (ii) Iago *de bedoeling heeft* Cassio *ervan te overtuigen* dat Desdemona *zal proberen* Othello *te doen inzien* dat Cassio *de bedoelding had* om het belang van de publieke zaak te dienen toen hij Montano aanviel.

Zo gesteld klinkt het onmogelijk ingewikkeld, maar tegelijk weten we dat dit toneelstuk al eeuwenlang door talloze verschillende lezers en theatergangers wordt begrepen en gewaardeerd. Wat doet de tekst precies in die eerste twee akten om te zorgen dat het publiek deze zesvoudige inbedding van perspectieven kan volgen, en kan meenemen in zijn verdere begrip van de ontwikkeling van het plot? Ik stel dat hier zes "strategieën van expositie" (*expository strategies*) bij komen kijken. Mijn analyse concentreert zich op de tekst en het verhaal dat hierin wordt overgebracht, en gaat niet in op de bijzonderheden van opvoeringen van het stuk door acteurs op het toneel.

Om te beginnen moeten er door de tekst personages worden opgeroepen aan wie het publiek gedachtentoestanden kan toeschrijven. Hierdoor ontstaat er een fictioneel netwerk van personages die interacties met elkaar hebben, en die van alles over elkaar zeggen en denken. De gebeurtenissen van het verhaal worden beleefd, waargenomen en verteld via de perspectieven van leden van dit netwerk. Het effect is dat het publiek het ene moment met het perspectief van het ene personage meekijkt, en het andere moment met dat van het andere personage. Er is dus geen moment waarop er

"van buitenaf" moet worden gekeken naar het totaal van in elkaar ingebedde perspectieven, zoals dat gesuggereerd wordt in zin (ii) hierboven. Niettemin geeft zin (ii) een getrouwe, zij het wat kunstmatige, samenvatting van de situatie aan het einde van de tweede akte: het valt niet te ontkennen dat in *Othello* een complex netwerk van onderling samenhangende perspectieven centraal staat. Dus zelfs als het publiek ze dankzij de eerste twee strategieën geleidelijk en "van binnenuit" kan leren kennen, in plaats van inéén als hele stapel, blijft de vraag staan hoe deze complexiteit door de tekst behapbaar wordt gemaakt. Ik betoog dat het antwoord besloten ligt in de resterende vier strategieën. Ten eerste is dit *framing*, het principe dat door het aanduiden van één bepaald concept een heel "raamwerk" (*frame*) van achtergrondkennis wordt geactiveerd bij het publiek. Met andere woorden: er hoeft niet bij nul te worden begonnen bij het overbrengen van hoe de onderlinge verhoudingen tussen de personages liggen. Het publiek is al bekend met begrippen als "wraak", "complot", "vriendschap" en "overspel", en kent de scenario's en rollen die hierbij passen. De tekst bouwt voort op algemeen gedeeld veronderstelde kennis van deze *frames* en schuift ze op ingenieuze wijze in elkaar. Zo ontstaat met weinig woorden de situatie die de kern van het stuk uitmaakt: Iago wil wraak nemen en zet een complot op waarin hij probeert Othello wijs te maken dat zijn vrouw overspel pleegt met Cassio, terwijl ze eigenlijk alleen vrienden met hem is. Het publiek heeft hiermee al een hele serie verwachtingen klaar over wie wat zal denken, hopen, bedoelen, enzovoorts. In plaats van dit allemaal stap voor stap te hoeven uitleggen kan een auteur volstaan met het "managen" van deze verwachtingen via een paar welgekozen details en plotwendingen—en dat is ook precies wat in de tekst van *Othello* terug te zien is.

De vierde en vijfde strategie hebben betrekking op de verhouding tussen het tijdsverloop van het vertelde verhaal (ook wel *fabel* of *story* genoemd) en dat van de vertelling (ook wel *sujet* of *plot*). In de narratologie is veel onderzoek gedaan naar mogelijkheden om hier flexibel mee om te springen (vooruitwijzingen en *flashbacks* zijn hiervan de bekendste voorbeelden). In *Othello* wordt hiervan handig gebruik gemaakt bij de presentatie van hoe perspectieven van verschillende personages zich tot elkaar verhouden. Het publiek krijgt het verslag voorgeschoteld van gebeurtenissen die zich ontwikkelen in minder dan 48 uur, terwijl er op de achtergrond verhaallijnen

spelen die veel meer tijd in beslag nemen. Deze worden in gecomprimeerde vorm aan het publiek meegegeven om, waar nodig, de verhoudingen te verduidelijken of de geloofwaardigheid te vergroten. Bovendien vindt de presentatie van de belangrijkste gebeurtenissen en plotwendingen *episodisch* plaats. Eerst onderhandelen twee of drie personages iets uit. Het resultaat hiervan wordt vervolgens weer meegenomen naar een volgende scene, waar opnieuw twee of drie personages tot een volgend resultaat komen, enzovoorts. Zo worden er nooit te grote stappen in een keer genomen, zoals dit wel het geval is als je zonder verdere context zin (ii) krijgt voorgeschoteld. De zesde strategie, tot slot, is eenvoudig: redundantie. Teksten zoals *Othello* herhalen belangrijke informatie gerust veelvoudig. Dankzij rijke stilistische variatie en doordat dit telkens vanuit verschillende perspectieven gebeurt, valt dat nauwelijks op. Toch helpt het natuurlijk voor het begrip van complexe situaties als een stuk informatie maar liefst zes keer wordt verteld, zoals het geval is in een van de voorbeelden die ik bespreek.

De verankering van de zes strategieën van expositie die ik onderscheid in de langlopende traditie van de narratologie, de studie van universele kenmerken van verhalen, rechtvaardigt de conclusie dat niet alleen *Othello*, maar verhalen meer in het algemeen in staat zijn om complexe verhoudingen tussen verschillende perspectieven inzichtelijk te maken. Uiteindelijk stel ik dat lexicale en grammaticale processen beneden zinsniveau geschikt zijn voor het onderling coördineren van ten hoogste twee tot drie perspectieven, maar dat daarboven de verhalende modus het overneemt. In Hoofdstuk 5 werk ik deze gedachte verder uit in een theoretisch model van menselijke interactie, waarvan de hoofdgedachte als volgt is. Met taal kun je iemand anders' aandacht vestigen op een object, gebeurtenis of concept. Stel ik zeg "die voetballer" nadat mijn gesprekspartner en ik het net over de nieuwe aanwinst van een zekere voetbalclub hebben gehad, dan laat ik hem de aandacht richten op die voetballer. Door te zeggen "die middelmatige voetballer", vestig ik niet alleen de aandacht van mijn gesprekspartner op die voetballer, maar geef ik bovendien een bepaalde kijk/bepaald gezichtspunt mee, in dit geval een niet al te positieve evaluatie. Dit is iets wat je als taalgebruiker eigenlijk voortdurend doet: je brengt zaken onder andermans aandacht en onderhandelt gezichtspunten op die zaken uit. In dagelijks taalgebruik zijn dit doorgaans

jouw eigen kijk en die van je gesprekspartner. Ook is het mogelijk om gezichtspunten erbij te betrekken van derde personen die al dan niet aanwezig zijn op het moment van spreken. Maak ik er bijvoorbeeld van "die figuur van wie Piet zegt dat hij een voetballer is" , dan geef ik mijn eigen gezichtspunt aan (ik impliceer dat ik eigenlijk vind dat "die figuur" de kwalificatie "voetballer" niet waard is), plus het gezichtspunt van een derde partij, namelijk van Piet. Door gebruik te maken van indirecte rede ("Piet zegt dat…") haal ik als het ware Piets gezichtspunt even binnen in het proces van afstemmen van mijn eigen kijk met die van mijn gesprekspartner. In theorie zou ik nog meer perspectieven kunnen invlechten door bijvoorbeeld te zeggen "die figuur van wie Piet zegt dat zijn trainer vindt dat hij een voetballer is". In de praktijk is dit echter een zin die je niet zo snel zult tegenkomen: zodra er meer dan twee of drie perspectieven een rol gaan spelen, zie je dat taalgebruikers overstappen naar de verhalende modus. Veel waarschijnlijker is dus dat ik zoiets zou zeggen als "Ik sprak Piet gisteren en die vertelde dat hij met zijn trainer heeft gesproken. Die is van mening…(enzovoorts)".

Het is een onderscheidend kenmerk van narratief taalgebruik dat hierbij altijd andere gezichtspunten betrokken zijn dan die van spreker en geadresseerde. Immers, zelfs als ik een verhaaltje begin te vertellen over hoe ik gisteren op straat liep, dan creëer ik als het ware een personage: een eerdere versie van mijzelf, waarvan mijn gesprekspartner en ik het gezichtspunt binnenhalen in het afstemmen van onze gedachten. Waar grammaticale en lexicale elementen op zinsniveau nog geschikt zijn voor het afstemmen van twee of drie perspectieven en de verhoudingen daartussen, beschikt verhalend taalgebruik over een gevarieerde "gereedschapskist" om een veelheid aan perspectieven en gedachtentoestanden weer te geven en op een natuurlijke manier met elkaar te verbinden: de zes strategieën van expositie. Daarmee faciliteren verhalen in elk geval *communicatie*: wie een situatie wil delen waarbij meer dan twee à drie perspectieven een rol spelen, schakelt doorgaans over naar een narratieve modus. Hiermee lijken verhalen bovendien *cognitie* te ondersteunen: een veelheid van perspectieven die in een paar zinnen opgesomd, of in een tabel of figuur weergegeven, volstrekt onbehapbaar lijken, kunnen gevat in de vorm van een verhaal vaak gemakkelijk worden begrepen. De narratieve modus is zo niet alleen een manier van spreken, maar ook van

denken. Het leren beheersen ervan levert ons een uitbreiding van ons denkrepertoire op in precies het domein waar gedachtenlezen over gaat: het begrijpen van sociale situaties in termen van onderliggende perspectieven en gedachtentoestanden van de mensen waar we mee omgaan. Door narratieve strategieën voor perspectiefcoördinatie te internaliseren, verwerf je naar alle waarschijnlijkheid denkpatronen die je ook in de interactie met de werkelijke sociale wereld om je heen kunt gebruiken.

In Hoofdstuk 3 neem ik de inzichten van de analyse tot zover mee naar het discours van de roman. Romans zijn niet zomaar verhalende teksten: ze hebben een specifieke vorm waaraan verschillende conventies ten grondslag liggen. Een belangrijk deel hiervan heeft betrekking op de *verteller*, de persoon of instantie die het verhaal aan de lezer doorgeeft en hierbij keuzes maakt wat betreft stijl, waarop de aandacht gevestigd wordt, wat wanneer wordt verteld, enzovoorts. Opnieuw neem ik als uitgangspunt enkele passages uit teksten waarin een rijk "gedachtenlandschap" (*thoughtscape*) wordt weergegeven, met andere woorden: die een complex netwerk van perspectieven neerzetten. Ik ga de discussie aan met de momenteel breedgedragen opvatting van Lisa Zunshine (o.a. 2012) dat literatuur ons uitdaagt en prikkelt doordat het de grenzen van onze mogelijkheden als gedachtenlezers opzoekt. Zo suggereert zij dat Virginia Woolf's *Mrs Dalloway* (1925) een bij uitstek literair werk is, aangezien lezers voortdurend de (volgens haar) cognitief belastende taak moeten uitvoeren van het in elkaar inbedden van perspectieven tot aan de vijfde of zesde orde. Door middel van taalkundige en narratologische analyse laat ik zien dat cruciale passages in de tekst dit hun lezers juist helemaal niet laten doen. De verteller neemt de lezer mee op een soort "360-graden-rondleiding" door het gedachtenlandschap, het netwerk van onderling verbonden perspectieven dat de tekst zorgvuldig en geleidelijk aan heeft geconstrueerd. Ik stel dat dergelijke romans ons niet zozeer fascineren en prikkelen omdat we zoveel moeite moeten doen voor het doordringen tot de perspectieven van de personages, maar juist omdat we een inkijkje krijgen in complexe sociale situaties vanuit verschillende invalshoeken, op een manier die in het "echte" leven niet mogelijk is. De verteller doet het leeuwendeel van het gedachtenlezen voor ons, zodat wij als lezers lui achterover kunnen zakken en van het gedachtenlandschap kunnen genieten.

Hoofdstuk 4 belicht de zaken nog eens vanuit het beginsel van cognitieve en communicatieve economie. Het centrale voorbeeld is de nieuwsverslaglegging van de zogenaamde Pistorius-zaak. In de nacht na Valentijnsdag 2012 schoot de Zuid-Afrikaanse atleet Oscar Pistorius zijn vriendin Reeva Steenkamp dood in zijn huis. Hij beweerde dat het een tragisch ongeval was, maar de politie arresteerde hem op verdenking van moord. Journalisten moesten de ochtend na de schietpartij aan hun lezers overbrengen dat nieuwswebsite *Beeld.com beweerde* dat een woordvoerder *zei* dat de politie-inspecteur *stelde* dat Pistorius *claimt* dat hij *dacht* dat zij een inbreker was, terwijl de politie *denkt* dat hij *wist* dat hij op zijn vriendin schoot. Dit deden zij natuurlijk niet met dergelijke gecompliceerde zinnen. Wat de krantenkoppen en nieuwsstukjes van die ochtend laten zien is een veelvuldig gebruik van woorden die ik *viewpoint packages* genoemd heb, zoals bijvoorbeeld *alleged(ly)*, *accidental(ly)* of *mistaken(ly)*. Wat deze woorden gemeen hebben is dat ze allemaal impliciet een of meerdere perspectieflagen coördineren, die als het ware in hun betekenis zitten verpakt of "opgeklapt". Wie zegt dat "*allegedly* Y happened *accidentally*", zegt eigenlijk dat een onbekende partij claimt dat een persoon A niet wist dat actie X uitkomst Y zou hebben, terwijl wel bekend is dat Y het geval is (in het Nederlands zouden we zeggen "dat iets *per ongeluk* gebeurd *zou zijn*"; waar in het Engels de lexicale eenheid *allegedly* beschikbaar is, gebruiken wij op deze plaats doorgaans een constructie met *zouden*).

Een spreker of auteur kan ervoor kiezen om een situatie waarin meerdere gezichtspunten een rol spelen *analytisch* op te bouwen. Dat wil zeggen dat alle gezichtspunten afzonderlijk worden geïntroduceerd en onderling gecoördineerd, veelal met talige constructies zoals de directe en (vrije) indirecte rede (A *wil* dat B *denkt* dat X, terwijl A *weet* dat Y…enz.). Ook kan ervoor worden gekozen om deze situatie *holistisch* weer te geven in taal, door het gebruik van een pakketje (A *misleidt* B). Het is evident dat het gebruik van pakketjes om complexe situaties holistisch weer te geven de communicatieve economie ten goede komt: het is ideaal voor een krantenkop als een complexe situatie in een paar woorden kan worden gevat. De suggestie die ik doe aan het eind van Hoofdstuk 4, voortbouwend op Daniel Dennetts idee van "gereedschap voor de geest" ("thinking tools"), is dat het leren gebruiken van dergelijke pakketjes in de loop van de taalverwerving bovendien

cognitieve efficiëntievoordelen met zich meebrengt. Met andere woorden: kinderen die leren wat een *misverstand* is, wat *per ongeluk* betekent, of wat *misleiden* inhoudt, voegen hiermee ook scenario's toe aan hun begripsvermogen die als gereedschap functioneren bij het nadenken over complexe situaties waarin meerdere perspectieven een rol spelen.

In Hoofdstuk 6 komen veel verschillende lijnen samen. Hier bespreek ik de praktijk van het onderzoek naar gedachtenlezen zoals dat plaatsvindt via psychologische tests en in neurowetenschappelijke laboratoria. In een veelgebruikte methode krijgen proefpersonen eerst een verhaaltje voorgelegd waarin meerdere personages en hun perspectieven een rol spelen. Vervolgens moeten de deelnemers vragen beantwoorden van het type: "Klopt het dat A van plan was B ervan te overtuigen dat C dacht dat X het geval was?". De vragen variëren in complexiteit, gemeten in het aantal perspectieven dat ze omvatten. Zoals gezegd heerst er consensus dat er een bovengrens zit aan het aantal ingebedde ordes dat mensen gemiddeld aankunnen. Uitgaande van de meeste experimenten ligt deze grens rond vijf ordes, maar sommige onderzoekers beweren dat hij hoger ligt, zo rond de zeven. In een kritische analyse van de data en resultaten van een viertal experimenten breng ik kritiekpunten naar voren op de heersende orde in het onderzoek naar gedachtenlezen, gebaseerd op mijn bevindingen uit eerdere hoofdstukken. Om te beginnen gebruiken veel tests verhaaltjes voor het presenteren van de situatie, maar zinnen voor het stellen van de vragen. Gezien mijn bevinding dat in natuurlijk taalgebruik de verhalende modus het overneemt vanaf ongeveer drie ordes, ligt de gevonden grens van vijf of zeven ordes wellicht eerder in de manier van vragen besloten dan in het vermogen van deelnemers als zodanig om complexe sociale situaties te doorzien—immers, zinnen waarin vier of meer perspectieven voorkomen zijn erg onnatuurlijk en lastig te verwerken (zie bijvoorbeeld zin (ii) over *Othello* hierboven). Andere kritiekpunten hebben betrekking op de manier van tellen van complexiteit via het aantal zinsinbeddingen. Er komen in de verhaaltjes en vragen ook pakketjes en andere vormen van impliciete perspectiefcoördinatie voor, waardoor "opgeklapte" gezichtspunten ten onrechte niet worden meegeteld. Ook wordt de meervoudige complexiteit van het gedachtenlandschap dat de verhaaltjes weergeven hierdoor gereduceerd tot uitsluitend inbedding, terwijl gezichtspunten vaak een netwerk van op allerlei manieren

verbonden perspectieven vormen. Verder brengt het gebruik van ja/nee-vragen problemen met zich mee waarmee in een deel van de tests niet juist is omgegaan, en zijn er problemen aan te wijzen die te maken hebben met rechtstreeks in de verhaaltjes vertolkte gezichtspunten tegenover gezichtspunten die alleen langs indirecte weg zijn af te leiden.

Uiteindelijk ontstaat er een dilemma. Enerzijds vertoont het bestaande paradigma van gedachtenleesexperimenten en de hieraan verbonden conclusies allerlei problemen, vooral wanneer het bekeken wordt in het licht van mijn eerste vijf hoofdstukken over hoe taal en verhalen in de praktijk met multi-perspectiefsituaties omgaan. Anderzijds vertoont dit paradigma een aantal onmiskenbare consistenties. Zo is er bijvoorbeeld een vaak gerepliceerde correlatie tussen de score die iemand behaalt voor gedachtenleestestjes en verschillende indicatoren voor diens algehele sociale vaardigheid. Wie hoge testscores haalt, heeft bijvoorbeeld vaak ook een groter sociaal netwerk en een hogere hoeveelheid grijze massa in hersengebieden die worden geassocieerd met sociaal gedrag. Ik sluit Hoofdstuk 6 af met een bespreking van hoe dergelijke correlaties kunnen worden verenigd met mijn bevindingen en kritiek.

De collectieve, culturele dimensie van het sociale brein

In de Conclusie formuleer ik een serie beweringen, observaties en aanbevelingen voor toekomstig onderzoek, elk rustend op een gedeelte van de zes hoofdstukken. Allereerst stel ik vast dat perspectieven waar we in sociale omgang en communicatie mee te maken krijgen eerder een *netwerkstructuur* vertonen dan een structuur van recursieve inbedding. Zo'n netwerk heb ik een *gedachtenlandschap* (thoughtscape) genoemd. In de praktijk van taalgebruik in verschillende genres (toneelstukken, romans, journalistiek, gesprekken) lijkt te gelden dat het werk van het weergeven en onderling coördineren van perspectieven verdeeld is over een verscheidenheid van lexicale, grammaticale en narratieve betekeniseenheden. Weergave van meerdere perspectieven met gebruik van uitsluitend zinsinbedding (A *denkt* dat B *vindt* dat C *wil*...enz.) leidt al snel tot een gebrekkige weergave van het gedachtenlandschap: enerzijds zijn zulke zinnen ondoorzichtig en lastig te begrijpen, anderzijds geven ze een

misrepresentatie van de vaak subtiele complexiteit van het netwerk van perspectieven.

Uiteindelijk stel ik dat ons sociale brein een individuele en een collectieve, sociaal-culturele dimensie heeft. Kort gezegd heeft onze evolutie in een complexe sociale omgeving volgens mij niet alleen druk uitgeoefend op het groter en krachtiger worden van individuele breinen (de klassieke *social brain hypothesis*), maar ook op het vinden van cultureel overdraagbare oplossingen om deze breinen zo efficiënt mogelijk te gebruiken. Deze oplossingen zien we terug in de "gereedschapskist" die taal en verhalen ons bieden voor het omgaan met complexe sociale situaties en de daaronder gelegen gedachtenlandschappen. De lexicale, grammaticale en narratieve conventies van talen zijn het resultaat van generaties mensen die probeerden hun gedachten met elkaar en met die van derden af te stemmen. Hierdoor is er een schat aan "interactie-ervaring" gestold in deze conventies. Kinderen die opgroeien en hun moedertaal verwerven kunnen hierdoor "op de schouders van reuzen staan". Door zich de lexicale, grammaticale en narratieve conventies van een taal eigen te maken, nemen ze het totaal van interactie-ervaring dat in de culturele gemeenschap beschikbaar is op in hun individuele cognitieve systeem. Vergelijk ons sociale brein met een iPad: in de loop van onze evolutie is de hardware beter en krachtiger geworden, maar uiteindelijk bepaalt ook de software die erop geïnstalleerd is hoe snel, effectief en gemakkelijk we hem kunnen gebruiken voor het uitvoeren van bepaalde taken. Uiteraard gaat dit om het uitbreiden van mogelijkheden en prestaties binnen de grenzen van wat de hardware kan—er zijn geen apps waardoor je een iPad kunt laten vliegen. Hetzelfde geldt voor gedachtengereedschappen zoals de *viewpoint packages* (bijvoorbeeld *accidentally* of *per ongeluk*) die werden geactiveerd om de Pistorius-zaak te begrijpen: deze pakketjes zijn als het ware stukjes software die lezers in de loop van hun leven hebben geïnstalleerd, waardoor het nu mogelijk is om binnen de gegeven beperkingen van het brein (zoals werkgeheugen) op een snelle en economische wijze complexe informatie te verwerken.

Deze kijk op de zaken verklaart ook waarom we moeite hebben met het begrijpen van een zin als die in (ii) hierboven, terwijl we als we Shakespeare's *Othello* lezen of in het theater zien, met gemak kunnen volgen hoe alle zes of zeven perspectieven zich tot elkaar verhouden. De limiet van vijf ordes (of

zeven) die onderzoekers claimen te hebben gevonden via gedachtenlees-experimenten in het lab ligt niet in onze individuele breinen besloten, maar is het gevolg van het samenspel tussen de mogelijkheden en beperkingen van ons brein en de gebruikte vormen van representatie via taal en/of verhalen—de cultureel overdraagbare gereedschappen. In de praktijk van onze interacties hebben we dus een gedeelde set van deze gereedschappen beschikbaar voor het omgaan met de complexiteit die verschillende perspectieven met zich meebrengen. Worden die op de juiste manier ingezet, dan nemen ze een deel van het denkwerk van ons over. Niet alleen zorgen ze daarmee dat we niet zo snel tegen een limiet aanlopen, maar ook zorgen ze voor communicatieve en cognitieve economie: op het individuele niveau kunnen wij ons het grootste deel van de tijd veroorloven om "luie" gedachtenlezers te zijn. Het is niet de standaard (zoals gesuggereerd door Scott-Phillips en Sperber, zie boven), maar de *uitzondering* om precies uit te pluizen "wie-wanneer-wat dacht". Of, in termen van Hoofdstuk 4: veel perspectieven blijven opgeklapt in holistische pakketjes zitten, en alleen als het nodig is construeren we ze analytisch. Bijvoorbeeld pas als er een misverstand optreedt, gaan we ons afvragen of A inderdaad *bedoelde* dat B zou *denken* dat A *wilde* dat B *dacht*…

En daarmee zijn we terug bij het voorbeeld van mij en de buurman. In de loop van ons leven begeven we ons in allerlei verschillende culturele gemeenschappen en subgemeenschappen. Voorbeelden zijn Nederlanders, studenten aan de Universiteit Leiden, supporters van een bepaalde voetbalclub, liefhebbers van een bepaald muziekgenre, volgers van een bepaald televisieprogramma, enzovoorts. In elk van die gemeenschappen doet een bepaalde set van taal- en verhaalconventies de ronde. Als je geleidelijk aan deel gaat uitmaken van zo'n culturele (sub)gemeenschap, wordt er als het ware een kopie van deze set in ons geheugen geïnstalleerd. Binnen dezelfde culturele gemeenschap kunnen we er dus van uitgaan dat iedereen min of meer dezelfde set van culturele gereedschappen aan boord heeft.

Inderdaad blijkt uit mijn onderzoek dat dit precies is wat mensen normaal gesproken doen tijdens een heleboel interactiemomenten. Ze handelen bij voorbaat al *alsof alle achtergrondkennis gedeeld is*. Met andere woorden, in plaats van meteen al te redeneren over wat A denkt dat B bedoelt dat A wil…(enzovoorts), zetten ze eerst gewoon een stap. Pas als dit niet het

gewenste effect heeft of er een misverstand ontstaat, komen ingewikkelde gedachtenleestaken in beeld. Wat wilde de buurman dat ik begreep dat hij bedoelde? We deelden niet dezelfde relevante set van verhaaltjes—ik wist niets van zijn modepolitie-act af. Als we wel dezelfde set hadden gedeeld, wat in de meerderheid van de interacties die we de hele dag door hebben het geval is, dan was de hele redenering over dat hij *bedoelt* dat ik *begrijp* dat hij *wil* dat ik *denk*... buiten beeld gebleven. Dan was het moment voorbijgegaan zonder dat ik extra energie in gedachtenlezen had hoeven steken—en die extra energie, die bewaart de luie gedachtenlezer alleen voor als het echt nodig is.

Acknowledgements

I met Ineke Sluiter and Arie Verhagen independently during my studies at Leiden. However, in 2007, both invited me to the same event: a three-day workshop on literature and evolutionary theory they were co-organising. At this event we discovered a joint interest: investigating what the study of language and literature can contribute to our understanding of the human mind and its evolution. Ever since that time I have been studying and working together with them on various occasions. As my supervisors, Arie and Ineke have done more than any student could ever deserve or wish for. During countless meetings they have been great "sparring partners", always prepared to listen to my findings, read my drafts, and challenge my ideas. I am extremely grateful that they have always kept on fostering my development as a scholar with such great enthusiasm and dedication. The relationship we have developed over the years is of much importance to me, intellectually as well as personally.

Studying at Leiden University felt good from the start, and so did working there. Rick Honings has been my friend since freshman's year and has always made it clear that it was out of the question that either of us would leave Leiden without a PhD. What a pleasure to have him standing by my side during my defence. I owe a great deal of gratitude to my friends and former office mates Barend Beekhuizen and Bob Corthals, whose support in academic as well as personal matters cannot be overstated. Joining the LUCAS PhD Council together with Sara Polak, Anna Dlabacova, Jacqueline Hielkema, David Louwrier, and Odile Bodde provided the perfect start at the institute, and I found it valuable to have them as my peers over the past years. I have good memories of the times spent with my colleagues from LUCAS and LUCL. I will never forget the party gig with the LUCAS Band formed together with Jenny Weston, Tazuko van Berkel, and Frans Willem Korsten.

I want to thank my colleagues from "de gang" at the Classics department for welcoming me as an outsider in their midst. Lia ten Brink, Viola Stoop, and Korrie Korevaart from LUCAS have always done a lot to help me out and support my activities, and so have the members of the LUCAS board, Kitty

Zijlmans, Frans-Willem Korsten, Thony Visser, and Geert Warnar. I am also grateful to and very much enjoyed working with the staff and students at the Dutch department.

I have been a regular at conferences of the European Human Behaviour and Evolution Association (EHBEA) since 2009. Not only have I learned a great deal there, but I have also met a number of amazing people over the years. I want to thank Anna Rotkirch and Tamás Dávid-Barrett, with whom I have the privilege of collaborating, for their great support and belief in me. Robin Dunbar has been a fantastic teacher and a generous host during my period at Oxford, and being part of his uniquely multidisciplinary research group (SENRG) is among the best working experiences I have ever had. Along with Robin Dunbar, I want to thank Felix Budelmann and Laurie Maguire for the pleasant collaboration while organising the Drama and Cognition lecture series at TORCH. Furthermore, Oxford certainly holds the highest concentration of clever and interesting people in the world, and living there was a constant source of joy. Many thanks for the great times go out to Bronwyn Tarr, Rafael Wlodarski, James Carney, Jacques Launay, Ellie Pearce, Isabel Behncke, Beau Lotto, Nathan Oesch, Wouter Wolf, Rogier Mars, Lennart Verhagen, Kantik Ghosh, Lauren Heathcote, Claudia Wehrspaun, and David Zollman. I owe a lot to Renée Hirschon, who was my landlady as well as a dear companion and teacher of the Oxford ways. The Magdalen College community and staff, and the MCR in particular, have been most welcoming and supportive of my social, sportive, and otherwise para-academic life in Oxford. The Leiden University Fund (LUF), LUCAS, and the NWO Spinoza Prize awarded to Ineke Sluiter have made my stay abroad possible with generous donations.

I thank Bob Corthals and Rob Zandvliet for their insightful comments on earlier versions of parts of my manuscript. Early on in the project, Irene de Jong and Barbara Dancygier have helped me find my way in the literature on viewpoint and fictional minds, which has been of great use. The outlines of Chapter 3 have taken shape during James Carney's and Emily Troscianko's cognitive humanities writing club, and Chapter 6 has benefited from the discussion at the Mentalising Workshop organised by Jacques Launay, Robin Dunbar, and me at Magdalen College in September 2014. I thank all who contributed to these occasions. I am also grateful to the Praemium Erasmianum 2012 and Faces of Science teams for helping me speak and write about my

research for a broad audience, which helped me organise my thoughts. All in all, both the current thesis and my personal development as a researcher have benefited greatly from discussions with the many bright and skilled people I have been surrounded by during the past years. Apart from those whom I have mentioned they include, in no particular order: Ton Anbeek, Jaap Goedebuure, Kimmie Jautze, Bas Haring, Clementine Breedveld, Louise Barrett, Andrea Macrae, Terrence Cave, Celia Heyes, Anneke Haddad, Victor Mids, Cathleen O'Grady, Philip Huff, Frans Saris, Steije Hofhuis, and John Neubauer. Thank you for letting me make use of your minds.

LUMC star doctors Anthonie Taminiau, Sander Dijkstra, Nicolette Leyerzapf, Michiel van de Sande, Hans Gelderblom, and Judith Kroep, along with Henk and Jet Hueting, not only secured that I could write this thesis at all, but also at times delivered important mental support over the past years. Be it never forgotten what they have done.

Finally, I have always been extremely blessed with my family and friends, and this has become clearer to me than ever during the past few years. Some I have met during my PhD, others go back many years or even decennia. John has listened carefully to my thoughts and worries whenever we spoke. My cordial Status is going strong as always after 13 years. My band mates Arjan, Martijn, Maarten, Thijs, and Thomas have taken care of drawing me out of academic thinking whenever needed. There is a wonderful piece of material evidence of Cees' belief in my doctorate dating 10 years back. Daniël, my brother in all respects but birth, will cover my other flank during the defence. My sister Ellen has been a wonderful friend and firm support all my life, and no less during the past four years. Her beautiful son Bram and husband Justin guarantee an indispensible dose of humour every time I see them.

I dedicate this thesis to my parents, Aad and Mia, who, from my first day on this planet have done everything and more to foster my curiosity and support my interests and plans. They have prepared me for this life with insight and knowledge, and, most importantly, by teaching me how to make it worth living for myself and others around me. Having found an occupation, with their support, that suits me so well is of the greatest value. However, the main reason why doing this PhD was the best thing that ever happened to me is because I met Odile, being with whom is what makes me sleep well at night, and what gets me up in the morning.

Curriculum Vitae

Max Johannes van Duijn (1984) studied Dutch Language and Literature at Leiden University (BA, *cum laude*) and completed minors in Business Administration and Philosophy. In 2008 he enrolled in the research master Literature and obtained permission to specialise in cognitive approaches and philosophy of mind by including courses at the Linguistics and Philosophy departments in his curriculum. He graduated in 2010 (MPhil, *cum laude*). In 2005 he started working part-time in marketing and IT business, which he kept on doing until taking up a full-time PhD position at the Leiden University Centre for Arts in Society (LUCAS) in 2011.

In 2012-2013 he was a member of the LUCAS PhD Council and got the opportunity to teach two undergraduate courses at the Dutch department. In 2014 he joined the Social and Evolutionary Neuroscience Research Group (SENRG) at the Department of Experimental Psychology, University of Oxford, as a visiting researcher. In Oxford he co-organised the Drama and Cognition lecture series at TORCH. In 2013 he co-initiated the Fraternity Friendship Study (FFS), a longitudinal investigation of social network formation and friendship in European student fraternities. After finishing the current thesis he started working as an Assistant Professor and postdoctoral researcher at Leiden University.

Since 2014 writes writing popular-science blogs for Faces of Science (KNAW) and frequently gives public lectures related to his research. He is a fervent sea sailor and plays the guitar in coverband Boozzz.