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

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

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.