Fiction and Common Ground
A Workspace Account

Merel Semeijn

rijksuniversiteit
groningen

Fiction and common ground

A workspace account

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

maandag 30 augustus 2021 om 9.00 uur

door

Merel Semeijn

geboren op 27 juni 1993
te Waddinxveen

Dedicated to Aletta Henriëtte Jacobs

Contents

Contents

Contents

1 Introduction

1.1 The goal

Generally, I trust people to tell me the truth. When I ask a stranger at a bus stop whether line 4 stops there, I expect the stranger to (unconsciously) follow Grice's (1989) maxim of Quality, i.e., my conversational partner will try to make their contribution to our conversation one that is true. This trust is essential; if a society lacks the trust that most people will engage in cooperative information exchange, the merit of its entire linguistic practice is in jeopardy. Why would I even try asking the stranger about the bus if I cannot trust that they will try to say something true?

At the same time people simply *love* to watch others break the maxim of Quality or even to break it themselves. We love reading or watching obviously false reports about unlikely romances in nineteenth century England, fire-breathing dragons, talking animals, heroic aliens and futuristic societies. Not only do we fill our bookcases, bedroom walls and minds with stories and images that we know to report false 'facts', we also love to extensively discuss these falsities. High schools oblige their students to rehearse and study their society's most famous false stories (e.g., *Hamlet* or *Van den Vos Reynaerde*) and people collectively create extensive online encyclopedias reporting on their favourite sets of falsities (e.g., *One Wiki to Rule Them All* or *Wookieepedia*).

People can only enjoy this blatant breaking of the maxim of Quality precisely because it is out in the open for everyone that the relevant stories are false. It is usually clear from the context that they are untrue and hence can be kept separate from other discourse. This is one of the hallmarks of our fiction telling practices (and what distinguishes it from its cousin 'lying' that, I hope, fewer people enjoy). Nobody is deceived into believing something false, nobody is harmed and we can maintain the general trust that we have in conversational partners.

1.1.1 The puzzles

This dissertation focuses on modelling the different ways we use language to engage with fiction and how these uses relate to non-fictional but related language uses such as assertion and lying. This general challenge subdivides itself into several puzzles involving different kinds of statements. The aim of the present dissertation is to engage in the existing debates on these puzzles by developing a coherent semantic analysis that deals with these distinct statements in a systematic way.

The central conundrum (and the starting point of the study of the semantics of fiction in general) is to distinguish regular non-fictional talk from fictional talk (see chapters 3 and 4). Whereas non-fictional discourse is typically about real world objects (e.g., Trump, the Eiffeltower, my aunt), fictional discourse typically features fictional names, i.e., names of fictional characters or objects such as 'Frodo', 'The Toothfairy' or 'the Fountain of Youth'. To illustrate, suppose Tolkien told me that his friend C.S. Lewis was born in Belfast by stating (1). Compare this to Tolkien's written statement (2) in his novel *The Hobbit*:

(1)   C.S. Lewis was born in Belfast.

(2)   In a hole in the ground there lived a hobbit.

By uttering (1) Tolkien engaged in non-fictional discourse, by writing down (2) Tolkien engages in fictional discourse. The aim here is to explain why, whereas statements such as (1) are considered to be true or false, statements such as (2) are not *really* true or false, but rather determine what is true in a fictional world separate from the actual world; (2) makes it true in the world of the *The Lord of the Rings* saga that a hobbit lived in a hole in the ground.

A complication here is that many things will be true in a fiction that were never actually stated in the fictional narrative. For instance, it is true in the world of *The Lord of the Rings* that water is H_2O, that the sun rises in the east and that handkerchiefs are squares of thin fabric used for personal hygiene. What's more, sometimes statements found in a fictional story do *not* determine what is true in the fictional world in this way. Consider Kesey's written statements in his novel *One Flew Over the Cuckoo's Nest*:

(3)   Then, just as she's rolling along at her biggest and meanest, Mc-Murphy steps out of the latrine door right in front of her, holding that towel around his hips – stops her *dead*! She shrinks to about

head-high to where that towel covers him, and he's grinning down on her.

Although these statements express that nurse Ratched (who previously grew to the size of a truck) shrinks to the size of a hobbit, they cannot be taken at face value. The narrator is schizophrenic and hence unreliable; it is not true in the world of *One Flew Over the Cuckoo's Nest* that Ratched shrinks in size. Rather it is probably true that she was taken aback and that the narrator perceived this as shrinking in size. Such cases of unreliable narration pose an additional puzzle for an account of fictional discourse (see chapter 8).

Moreover, sometimes people talk about fiction and use fictional names in statements that are not part of a fictional story. For instance, I may utter (4) or (5):

(4)   In *The Lord of the Rings*, Frodo was born in the Shire.

(5)   According to *The Lord of the Rings*, Frodo was born in the Shire.

(4) and (5), like fictional statement (2), tell us something about a fictional world. However, like the non-fictional statement (1), (4) and (5) seem to be really true! But do statements (4) and (5) really mean the same thing (see chapter 7)? To complicate things, I can also utter a statement such as (6):

(6)   Frodo was invented by Tolkien.

Arguably, (6) expresses something that is really true. But, assuming the name 'Frodo' refers uniformly to one object, how can we make sense of the apparent truth of both (4) *and* (6)? If Frodo is the kind of thing that was 'born' in a certain region, how can he have been 'invented' by someone (see chapter 6)?

Conversely, sometimes we may find a statement *within* a fictional story that is really true about non-fictional objects (see section 4.6.2). For instance, in Fleming's fictional novel *Thunderball* we come across the following statement:

(7)   New Providence, the island containing Nassau, the capital of the Bahamas, is a drab sandy slab of land fringed with some of the most beautiful beaches in the world. (Friend (2008))

(7) expresses information that is simply true. New Providence *does* indeed contain Nassau and is fringed with beaches. Does this then mean that it is not part of the fictional story? And if it is part of the story, how can we

explain that I may learn topographical facts about the Bahamas from reading (7) in the fictional novel *Thunderball*?

Finally, whenever one attempts to model our fiction-telling practices, it is important to keep fictional discourse separate from lies (see chapter 5). Suppose I am invited to a party that I do not want to attend. I could utter (8) while I am feeling perfectly fine and hence lie:

(8)   [Cough, cough] Sorry, I have a cold.

Although both lie (8) and fictional statement (2) are strictly false, we want to maintain a distinction between these two types of utterances; a semantic definition of fiction should not also apply to lies and vice versa. Intuitively, we may want to appeal to the notion of an intention to deceive; whereas I had an intention to deceive my addressee with my utterance of (8), Tolkien never intended to deceive anyone into believing that a hobbit lived in a hole by writing down (2). However, sometimes it seems like we can lie without any intention to deceive. Suppose a student accused of plagiarism is called to the dean's office. The student knows he plagiarized, the dean knows he did, the student knows that the dean knows, etc. However, it is also well known that the dean will not punish anyone who explicitly denies their guilt. When asked the student therefore says:

(9)   I didn't cheat on the exam. (adapted from Carson (2006))

If (9) is a real lie, then lies apparently do not necessarily involve an intention to deceive. But then how can we distinguish lying from fiction telling?

1.1.2 Theoretical ingredients

This dissertation introduces and explores a coherent semantic theory of the different types of statements described above called the 'workspace account'. It is an extension of Stalnaker's (1970, 1978, 1984, 2002) widely adopted pragmatic common ground framework. In this framework assertions are modelled as proposals to update the 'common ground' of a conversation (i.e., the set of propositions that are mutually presupposed by speaker and hearer). The Stalnakerian common ground framework is based on cooperative information exchanges: Conversations in which interlocutors gradually add more and more information to their common ground because more and more propositions are asserted and accepted. Since the workspace account is aimed at modelling fictional discourse (e.g., Tolkien's production

of his fiction novel *The Lord of the Rings*) and fiction authors do not, strictly speaking, engage in cooperative information exchange (e.g., the communicated information in *The Lord of the Rings* is strictly false), I will need to extend this basic Stalnakerian notion of common ground. In developing the present extension of the common ground framework I have taken inspiration from (amongst other things) Eckardt's (2014) and Stokke's (2013; 2018) 'unofficial common ground accounts' (that separate the common ground for non-fictional discourse from common grounds for fictional discourse), Matravers' (2014) theory of narrative interpretation (according to which our primary engagement with narratives – entertaining its content – is the same whether the narrative is fictional or not) and Lewis' (1978) widely adopted possible world analysis of statements of the form 'In story s, ϕ'.

The notion of common ground will be represented in different formal ways in the dissertation. Familiarity with basic set theory and propositional logic is assumed. Familiarity with Kamp's (1981) Discourse Representation Theory will be useful in reading the dissertation but not strictly necessary because the formalism is specified and explained where introduced. The DRT notation is the most frequently used notation in this dissertation because this formalism allows me to illustrate the different kinds of common ground updates and the anaphoric dependencies that are involved in the example sentences in a visually efficient manner.[1]

1.2 Overview of dissertation

1.2.1 Summary of each chapter

Chapter 2: Common ground: *In sensu composito* or *in sensu diviso*

In chapter 2 I introduce the notion of common ground and its role in Stalnaker's pragmatic framework. I briefly discuss previous challenges that have been posed to traditional common ground definitions and spell out a novel challenge posed by non face-to-face communication using a relational analysis of de re beliefs. Basically, standard common ground definitions are in terms of iterative de re attitudes but surely Tolkien never had any de re beliefs about me (one of his readers). But then how can we talk about the

---

[1]As will become clear there is also a theoretical incentive to opt for a formalism that gives a structured representation of the common ground. See chapter 4.

conversation or common ground between Tolkien and his readers? I propose and compare two potential refinements of standard common ground definitions based on Abelard's distinction between generality *in sensu composito* and *in sensu diviso*.

Chapter 3: Unofficial common grounds

In chapter 3 I turn to fiction. I introduce the basic puzzle in the semantics of fiction of modelling the quarantining of fictional discourse from non-fictional discourse. I discuss and formalise existing Stalnakerian accounts that model fictional discourse, i.e., Stokke's and Eckardt's unofficial common ground accounts where fictional statements update an unofficial common ground that is kept separate from the official common ground used for assertions. I critique both accounts on how they deal with two prima facie conflicting intuitions concerning fictional content. First, that fictional truths are only accepted temporarily. Second, that we do retain information about fictional truth even after engaging with the fiction somehow.

Chapter 4: The workspace account

In chapter 4 I introduce a novel Stalnakerian approach to model fictional and non-fictional discourse, i.e., the workspace account. I first discuss two theoretical ingredients of the account: Matravers' theory of fiction interpretation and Lewis' analysis of the fiction operator. I then discuss the basic ideas of the workspace account and formalise them. Assertions and fictional statements both uniformly update a temporary common ground (the workspace) during the ongoing discourse. What distinguishes the speech acts is how they update the stable common ground, i.e., with information of the form 'In story s, ϕ' or with plain ϕ. I will argue that the workspace account avoids the difficulties associated with the unofficial common grounds described in chapter 3. I briefly discuss two possible extensions of the workspace account: a version of the account where fiction that mentions real world entities is de re about these entities and an analysis of export of fictional truth as based on analogical reasoning with parafictional information.

The chapters that follow deal with possible applications and refinements of this basic account. Sometimes these adjustments can easily be translated to the unofficial common ground accounts. I will discuss this when relevant.

Chapter 5. Lies, bald-faced lies and parafictional updates

In chapter 5 I explore the challenges posed by modelling lying in a common ground framework and discuss two possible strategies (open to unofficial common ground accounts and the workspace account) to deal with these. I argue that bald-faced lies (i.e., statements that are called "lies" but that do not involve an intention to deceive) are best analysed as fictional statements rather than as actual lies. I briefly introduce and counter five objections to this view.

Chapter 6: The challenge of metafictional anaphora

In chapter 6 I introduce metafictional discourse (i.e., statements about fictional entities *as fictional entities*) and the different variants of the problem of the wrong kind of object, i.e., if Frodo is a flesh and blood hobbit, how can he also have been invented by Tolkien? I specify the challenge posed by the possibility of co-predication and anaphoric dependencies across metafictional and parafictional discourse for dynamic semantic approaches. I focus on pronominal anaphora across mixed parafictional/metafictional discourse or 'metafictional anaphora'. I evaluate four different possible solutions in the workspace account.

Chapter 7: The 'In' and 'According to' operators

In chapter 7 I argue, contra common practice, that there is a relevant semantic difference between the 'In story *s*' and 'According to story *s*' operators. I explore three novel observations concerning the diverging linguistic behaviour of these operators and propose two distinct semantic analyses to account for these. Sentences of the form 'In story *s*, ϕ' receive the widely adopted Lewisian analysis of parafictional statements. Sentences of the form 'According to *s*, ϕ' are analysed as indirect speech reports.

Chapter 8: Unreliable narration and imaginative resistance

In chapter 8 I discuss the need for some kind of revision operation for workspace updates and how this relates to the overall project of a Matraversian 'uniform' analysis of fictional and non-fictional discourse. The rest of chapter 8 is based on a co-authored paper with Dr. Emar Maier. We present an extension of the workspace account that takes into account the role of

(unreliable) narrators and the phenomenon of imaginative resistance by incorporating insights from belief revision.

1.2.2 How to read this dissertation

The most straightforward way of reading this dissertation is to read all chapters in consecutive order. This way the reader will obtain a comprehensive overview of the workspace account and its different applications. If the reader is interested in a particular topic they may choose to skip chapter 2 (and rely on an intuitive understanding of what it is to have a common ground with someone who is not acquainted with you) and start with chapters 3 and 4 to familiarize themselves with the basics of the workspace account and the motivations behind it. After reading chapters 3 and 4 the reader can move on to any of the subsequent chapters 5, 6, 7 or 8. The text may at some points include references to ideas, analyses or examples in earlier and later chapters (other than chapters 3 and 4) but these may be skipped. Chapter 7 may also be read completely independently if one has read section 4.2.2.

1.3 Previous publications

Most of the research presented in this dissertation is based on (parts of) papers that have been presented at conferences. Some research is based on previously published papers. Every chapter starts with a short description of what parts of the text are taken from what publication and what has been added or changed. Below is an overview of which papers formed the basis for which chapters:

Chapter 2: Common ground: *In sensu composito* or *in sensu diviso*
   Semeijn, M. (2019). Common ground: In sensu composito or in sensu diviso. In *Proceedings of 22nd Amsterdam Colloquium*. ILLC.

Chapter 3: Unofficial common grounds
   Semeijn, M. (2017). A Stalnakerian analysis of metafictive statements. In *Proceedings of the 21st Amsterdam Colloquium*. ILLC.

Chapter 4: The workspace account

Semeijn, M. (2017). A Stalnakerian analysis of metafictive statements. In *Proceedings of the 21st Amsterdam Colloquium*. ILLC.

Semeijn, M. (2019). Interacting with fictions: The role of pretend play in Theory of Mind acquisition. *Review of Philosophy and Psychology*, 10(1):113-132.

Chapter 5: Lies, bald-faced lies and parafictional updates

Semeijn, M. (2020) Bald-faced lies and parafictional updates. unpublished manuscript.

Chapter 6: The challenge of metafictional anaphora

Semeijn, M. (2019) The challenge of metafictional anaphora. In Sikos, J. and Pacuit, E., editors, *At the Intersection of Language, Logic, and Information*, 124-143. Springer, Berlin.

Semeijn, M., Zalta, E. N. (2021). Revisiting the 'wrong kind of object' problem. *Organon F*, 28(1):168-197.

Chapter 7: The 'In' and 'According to' operators

Semeijn, M. (2020). The 'In' and 'According to' operators. In *Proceedings of the ESSLLI & WeSSLLI Student Session 2020*.

Chapter 8: Unreliable narration and imaginative resistance

Maier, E., Semeijn, M. (forthcoming). Extracting fictional truth from unreliable sources. In Maier, E. and Stokke, A., editors, *The Language of Fiction*. Oxford University Press, Oxford.

2 Common ground: *In sensu composito* or *in sensu diviso*

This chapter is a rewritten and expanded version of 'Common ground: *In sensu composito* or *in sensu diviso*' in *Proceedings of 22nd Amsterdam Colloquium*. The most substantial differences between this chapter and the proceedings paper include: First, an expansion of the discussion of the notion of common ground and its traditional definitions (section 2.2.1). Second, the addition of a brief discussion of previous challenges concerning the inapplicability of these traditional definitions to cases of real-life communication (section 2.3.1). Third, the addition of a discussion of the different notions of felicity resulting from the proposed common ground definitions (section 2.5.1).

2.1 Introduction

The central aim of this dissertation is to model our engagement with fiction in Stalnaker's (1970, 1978, 1984, 2002) pragmatic framework, i.e., in terms of what is and what isn't 'common ground' between interlocutors. To prepare the way for the application of the notion of common ground to the case of fiction in subsequent chapters, I will first focus solely on exploring and refining this notion in the current chapter. Several challenges to traditional common ground definitions will first have to be addressed, including especially a novel challenge that I introduce concerning conversational participants that are not acquainted with each other. Traditional 'textbook' definitions of common ground are based on face-to-face conversations where people have iterative attitudes towards each other. For instance, we can define common ground as common belief: p is common ground between a and b iff a believes that p, b believes that p, a believes that b believes that p, b believes that a believes that p, etc. However, note that if a believes that b has

a certain belief, this means that *a* has a belief about her interlocutor *b*. This is problematic because the concept of common ground has – without much hesitation – been extended to non-face-to-face communication (e.g., Stokke (2013); Eckardt (2014); Semeijn (2017); Maier and Semeijn (forthcoming); Zucchi (forthcoming)) in which the speaker is known to the hearer but the hearer is not known to the speaker, such as books, broadcasted speeches or blogposts.[1] For instance, it is common ground between biographer Ray Monk and myself that Wittgenstein was Austrian. However, definitions of common ground in terms of attitudes towards one's interlocutor do not apply to this type of communication; I may have beliefs about Monk but obviously Monk does not have beliefs about me. Monk merely has beliefs about the beliefs of 'the reader(s)', whoever that may be. It would follow that there could never exist any common ground between Monk and his readership. This is unsatisfactory since Monk and myself *do* seem to respectively produce and interpret the biographical text against a background of shared assumptions.

Non-face-to-face conversations pose a challenge to anyone who wants to formulate an account of our engagement with fiction in terms of common ground. If you want your account of fiction to not only apply to face-to-face 'around the campfire' fiction telling, but also apply to more stereotypical cases of fiction telling such as Tolkien's fictional novel writing, you will have to assume a notion of common ground that can be extended to non-face-to-face communication. The central aim of this first chapter is thus to get this basic challenge posed by non-face-to-face communication out of the way so we can proceed in subsequent chapters to analyse our engagement with fiction in terms of common ground. I will restrict the discussion in this chapter to cases of non-fictional conversations. If the reader is primarily interested in semantics of fiction, they may continue with section 2.2, skip the rest of the current chapter and – for the rest of the dissertation – rely on an intuitive understanding of something being common ground between people that are not acquainted.

I first discuss the concept of common ground and its role in Stalnaker's pragmatic framework (section 2.2). Second, I briefly discuss three challenges that have previously been posed to traditional common ground definitions

---

[1]I use the terms 'speaker' and 'hearer' for, respectively, the person that produces some communicative act (e.g., by saying it or by writing it down) and the person that receives and interprets the communicative act (e.g., by hearing or reading it).

(Section 2.3.1). Next, I spell out a novel challenge to traditional common ground definitions posed by non-face-to-face communication using a relational analysis of de re belief that makes explicit the acquaintance relations involved in de re beliefs (Section 2.3.2). Next, I introduce Abelard's distinction between generality *in sensu composito* and *in sensu diviso* and, in line with this distinction, propose two potential refinements of common ground definitions (Section 2.4). I discuss how we obtain two essentially different notions of the felicity of speech acts on the two definitions and, finally, show how the case of an acquaintance that hasn't revealed themselves as a conversational participant may aid us in deciding between the two definitions (Section 2.5).

2.2 Common ground

2.2.1 Common ground and assertions

The notion of common ground not only is a key concept in pragmatics, but it (and related notions such as 'mutual belief', 'mutual knowledge' or 'joint attention') also plays an important role in other fields including philosophy (e.g., Friedell (1969); Lewis (1969); Schiffer (1972)), epistemic logic and computer science (e.g., Fagin et al. (1995); Meyer and Van der Hoek, (1995)) in explaining how agents can coordinate their actions.[2] For instance, we avoid multiple collisions in traffic because it is *common* belief between drivers that they should drive on the right side of the road; it's not enough if all drivers mutually believe that drivers should drive on the right. To coordinate action they must also believe that the other drivers believe this, etc.

Stalnaker uses the concept of common ground to model communicative practices in collaborative inquiry. More specifically, Stalnaker's framework models how people's assertions contribute to a growth of shared information. The common ground framework is thus a type of dynamic semantic framework. The basic idea of dynamic approaches is that the meaning of a statement is its 'context change potential', i.e., how it updates an information state called the conversational 'context'. Traditionally, there has been a debate in dynamic semantics over what this conversational context exactly

---

[2]See van Ditmarsch et al. (2009) for an overview of the history of concepts such as 'common knowledge' and 'common belief' and various definitions that have been offered.

is. On a 'psychologistic' interpretation of the context (e.g., Geurts (1999); Kamp (2015); Maier (2017)) it is an agent's mental state. Speech acts are thus defined in terms of how they update a specific agent's propositional attitudes. On a 'common ground' interpretation of the context such as Stalnaker's (see also e.g., Heim (1982); Groenendijk and Stokhof (1991); van der Sandt (1992).) it is a representation of the common ground between speaker and hearer. Speech acts are thus defined in terms of how they update the (interlocutor neutral) common ground (but see Hamm et al. (2006) for a conciliatory view).

   In Stalnaker's framework, the common ground of a certain conversation is the set of presuppositions mutually shared by the conversational participants. This notion has a dual function: First, assertions are analysed as proposals to update the common ground. Second, the notion of common ground is used to explain how a shared background can guide the production and interpretation of speech acts. For instance, suppose Mary makes the following assertion about an upcoming party:

 (10)   Nobody is coming to the party.

Firstly, through this assertion Mary proposes that it becomes common ground between her and her hearer (let's call him John) that nobody is coming to the party. In other words, she proposes to further limit the 'context set' (i.e., the set of possible worlds compatible with the common presuppositions) by removing possible worlds in which someone *does* come to the party. Secondly, Mary can felicitously say (10) (and her assertion can be properly understood by John) precisely *because* it was already common ground between Mary and John that there is a party before Mary's assertion. If it hadn't been, Mary's assertion would not have been pragmatically well-formed.

2.2.2 Traditional definitions of common ground

According to Stalnaker (2002), a prima facie intuitive and simple approximation of the concept of common ground is to define it as common belief. I dub this the 'belief-based' conception of common ground. Stalnaker defines common belief as follows:

> [A] proposition ϕ is common belief of a group of believers if and only if all in the group believe that ϕ, all believe that all believe it,

all believe that all believe that all believe it, etc. (Stalnaker, 2002, p.704)

This formulation is often copied by semanticists that work with Stalnaker's pragmatic framework (e.g., Stokke (2013); Zucchi (forthcoming)). However, strictly speaking the formulation is ambiguous between a de re and a de dicto reading: Do all believers believe de re of all believers that they believe that ϕ, or do all believers believe de dicto that "all believers believe that ϕ"?[3]

The difference in interpretation comes out when we formalise the notion of common ground. Standard textbook analyses of notions such as common belief and common knowledge are in terms of iterative de re attitudes that people have towards each other. For instance, proposition 2.5 in the *Stanford Encyclopedia of Philosophy* article on common knowledge (Vanderschraaf and Sillari (2014)) defines common knowledge in terms of a 'hierarchy' of statements of the form '*a* knows that *b* knows that *c* knows that ... ϕ' (see also e.g., Geurts (2020); Schiffer (1972, p.30)). Common ground (on a belief-based conception) can thus be formalised as follows: *p* is common ground between everyone in some community of conversational participants iff (where Cx means x is a conversational participant and $\mathbf{B}_x\phi$ means x believes that ϕ):

$$\forall x(Cx \to \mathbf{B}_x p)$$
$$\forall x \forall y((Cx \wedge Cy) \to \mathbf{B}_x \mathbf{B}_y p))$$
$$\forall x \forall y \forall z((Cx \wedge Cy \wedge Cz) \to \mathbf{B}_x \mathbf{B}_y \mathbf{B}_z p))$$
$$\vdots$$

In words, everyone in the community believes that *p*; everyone in the community believes of all members of the community that they believe that *p*; everyone in the community believes of all members of the community that they believe of all members of the community that they believe that *p*; etc. From this we can derive a hierarchy of iterative de re attitudes. For instance, if *a*, *b*, and *c* are the conversational participants of some community (i.e., *Ca*, *Cb* and *Cc*), *p* is common ground in the community iff *a*, *b* and *c* all believe that *p*; *a* believes that *b* believes that *p*; *a* believes that *c* believes that *p*; etc.[4]

---

[3] Arguably, there is even a third interpretation possible where all believers believe de re of the group of believers that it collectively believes that ϕ. I leave exploration of this option to further research.

[4] Here and in the rest of this chapter I am assuming the principle of positive introspection (see Rendsvig and Symons (2019) for an overview of different epistemic logics and their

In a similar vein, we can talk about common ground between speaker and hearer of a particular speech act in terms of iterative de re beliefs: p is common ground between speaker a and hearer b iff:

$$
\begin{array}{cc}
\mathbf{B}_a p & \mathbf{B}_b p \\
\mathbf{B}_b \mathbf{B}_a p & \mathbf{B}_a \mathbf{B}_b p \\
\mathbf{B}_a \mathbf{B}_b \mathbf{B}_a p & \mathbf{B}_b \mathbf{B}_a \mathbf{B}_b p \\
\vdots & \vdots
\end{array}
$$

In words, a and b both believe that p; b believes that a believes that p; a believes that b believes that p; etc.[5] So when Mary says (10) to John she proposes that it becomes the case that she believes that nobody is coming to the party; that John believes this; that she believes that John believes this; that John believes that she believes this; etc. Similarly, Mary's assertion (10) is felicitous because Mary believes that there is a party; John believes this; Mary believes that John believes this; John believes that Mary believes this; etc.

Alternatively, common ground can be defined in terms of common acceptance. This is the notion of common ground that Stalnaker (2002) eventually adopts (see also e.g., Stokke (2013, 2018)): p is common ground between a and b iff (where $\mathbf{A}_x p$ means x accepts that p):

$$
\begin{array}{cc}
\mathbf{A}_a p & \mathbf{A}_b p \\
\mathbf{B}_b \mathbf{A}_a p & \mathbf{B}_a \mathbf{A}_b p \\
\mathbf{B}_a \mathbf{B}_b \mathbf{A}_a p & \mathbf{B}_b \mathbf{B}_a \mathbf{A}_b p \\
\vdots & \vdots
\end{array}
$$

---

principles): If a believes that p, then a also believes that a believes that p, etc. In other words, our attitudes are transparent to ourselves.

[5] Some theorists prefer to give a precise finite definition of these seemingly infinite concepts such as common belief and common knowledge. One way of doing this is by means of a recursive definition. For example, let q represent the fact that some proposition p is common belief between a and b. q is true iff a and b believe both p and q. See e.g., Barwise (1988) for a comparison of such definitions and the iterative definition. Another way of doing this is to model knowledge and belief in Kripke models (i.e., in terms of accessibility relations between possible worlds) and define common knowledge through the transitive closure of the union of all individual agents' accessibility relations (see Fagin et al. (1995)). Roughly put, p is common knowledge between a and b iff p is true in all possible worlds that are accessible through some chain of a's and b's accessibility relations. I will ignore these issues and use infinite definitions in the following since these clearly capture the intuitive notion of what is necessary for something to be common ground.

In words, *a* and *b* both accept that *p*; *a* believes that *b* accepts that *p*; *b* believes that *a* accepts that *p*; etc. Stalnaker (2002) defines 'acceptance' as a propositional attitude that is similar to belief but doxastically neutral. I can for instance accept that Santa lives on the North Pole for the sake of conversation while I do not believe this. I dub this the 'acceptance-based' conception of common ground.

More recently, common ground has been defined as common commitment (e.g., Geurts (1997)): *p* is common ground between *a* and *b* iff (where $\mathbf{C}_{x,y}\phi$ means *x* is committed to *y* to act on ϕ being true):

$$
\begin{array}{cc}
\mathbf{C}_{a,b}p & \mathbf{C}_{b,a}p \\
\mathbf{C}_{b,a}\mathbf{C}_{a,b}p & \mathbf{C}_{a,b}\mathbf{C}_{b,a}p \\
\mathbf{C}_{a,b}\mathbf{C}_{b,a}\mathbf{C}_{a,b}p & \mathbf{C}_{b,a}\mathbf{C}_{a,b}\mathbf{C}_{b,a}p \\
\vdots & \vdots
\end{array}
$$

In words, *a* is committed to *b* to act on *p* being true; *b* is committed to *a* to act on *p* being true; *b* is committed to *a* to act on it being true that *a* is committed to *b* to act on *p* being true; *a* is committed to *b* to act on it being true that *b* is committed to *a* to act on *p* being true; etc.

2.3 Real-life communication

The above definitions of common ground abstract away from actual language use in several ways. This has spurred a general movement in philosophy and semantics of challenging and redefining traditional notions of common ground so as to make them applicable to real-life communication. In this section I first briefly discuss three challenges that have previously been posed to traditional definitions of common ground (and related notions such as common knowledge). These challenges are relevant to the current undertaking of applying the notion of common ground to our stereotypical engagement with fiction, e.g., reading a fictional novel. Second, I introduce a novel challenge – that is also relevant to this end – concerning the application of traditional common ground definitions to communication where conversational participants are not acquainted.

2.3.1 Steps towards a functional notion of common ground

No infinite iteration of attitudes

A central and widely discussed problem with the belief-based and acceptance-based definitions of common ground (and related notions such as common knowledge) is that they involve infinite iterations of propositional attitudes. Obviously, actual people don't form infinitely many beliefs about each other's mental state. They will usually not get beyond third or fourth order beliefs. So, given that the iteration will always be broken off at some point in real life, there can never actually exist any common ground between conversational participants. To avoid this issue we may choose to talk about 'partial common belief' or *p* being 'first-order belief', 'second-order belief', etc. depending on how many iterations of beliefs the conversational participants have formed.

In this dissertation I follow Lewis' suggestion that the infinite structure in common ground definitions represent "a chain of implications [that follow from our beliefs], not of steps in anyone's actual reasoning. Therefore there is nothing improper about its infinite length" (Lewis, 1969, p.53). For instance, the infinite iteration of beliefs in the belief-based definition of common ground represents the 'implicit' beliefs of the conversational participants (i.e., what they would believe if they were perfectly rational and had unlimited cognitive capacities) rather than their actual 'explicit' beliefs (see also Heal (1978); Clark and Marshall (1981); Barwise (1988)). Geurts' (2020) commitment-based definition of common ground adopts a similar 'normative' approach; commitments express norms for behaviour, not actual cognitive attitudes, and hence can iterate infinitely.

Graded beliefs

Another issue with traditional common ground definitions (that are formulated in terms of attitudes) is that they presuppose full-blown belief or acceptance. However, sometimes people can be more or less certain of their beliefs. People can have a less than full-blown first-order belief (e.g., I may believe that I locked the door but am not one hundred percent sure about this). Moreover, people can be more or less sure about higher-order beliefs. This can for instance occur when conversational participants are dealing with unreliable communication channels and are not sure whether a certain communicative act has been transmitted and hence whether their

interlocutor has updated their beliefs (see 'the coordinated attack problem', e.g., Akkoyunlu et al. (1975); Gray (1978); Cohen and Yemini (1979); Fagin et al. (1995)). Hence, in such situations there cannot exist common ground. A possible weakening of notions of common ground in terms of full-blown belief (or acceptance) is to talk about graded beliefs and hence graded common ground. For instance, we can use the notion of common p-belief where x p-believes some proposition ϕ iff x believes ϕ with at least a probability of p (see e.g., Brandenburger and Dekel (1987); Stinchcombe (1988); Monderer and Samet (1989)) In this dissertation I mostly abstract away from this issue; in most of the following discussion I assume full-blown belief in and acceptance of propositions and hence full-blown common ground. An exception to this is the discussion in chapter 8 where confidence in some information being common ground is expressed in terms of an entrenchment ordering on propositions.

No simultaneous attitudes or commitments

A less well-known issue with traditional definitions of common ground and related concepts such as common knowledge is that they require *simultaneous* attitudes or commitments. This has spurred a debate in logic of distributive systems on how to obtain common knowledge in systems where messages may be delayed (see e.g., Halpern and Moses (1990); Fagin et al. (1999); Panangaden and Taylor (1998)). However, the requirement of simultaneous attitudes or commitments has received very little attention in semantics. This is somewhat surprising because the requirement makes traditional common ground definitions inapplicable to many forms of real-life communication.

Reconsider the belief-based conception of common ground between speaker a and hearer b. According to this definition, for some proposition p to be common ground the speaker has to believe p (i.e., $\mathbf{B}_a p$). However, this can't simply mean that a believes p *at some point in time*. This would entail that we can merely establish that p is common ground simpliciter; depending on whether there exist points in time (past, present or future) at which a and b have the required attitudes, p is common ground or not. However, we want to use the concept of common ground as a dynamic notion, i.e., what is common ground can change over time. Hence we somehow want to be able to talk about p being common ground at a certain time t_1 because conversational participants have the relevant attitudes towards p at t_1 (and

not being common ground at some other time t_2 because conversational participants lack the relevant attitudes at t_2).

Prima facie, to arrive at such a notion we can – on a belief-based definition of common ground – assume the following time indexing: p is common ground at t_1 between speaker a and hearer b iff (where $\mathbf{B}_x^t \phi$ means x believes that ϕ at time t):

$$\mathbf{B}_a^{t_1} p \quad \mathbf{B}_b^{t_1} p$$
$$\mathbf{B}_b^{t_1} \mathbf{B}_a^{t_1} p \quad \mathbf{B}_a^{t_1} \mathbf{B}_b^{t_1} p$$
$$\mathbf{B}_a^{t_1} \mathbf{B}_b^{t_1} \mathbf{B}_a^{t_1} p \quad \mathbf{B}_b^{t_1} \mathbf{B}_a^{t_1} \mathbf{B}_b^{t_1} p$$
$$\vdots \quad \vdots$$

In words, a and b both believe at t_1 that p; a believes at t_1 that b believes at t_1 that p; etc. When either a or b stops having one of the relevant iterated beliefs at a later time t_2 (e.g., the speaker no longer believes that the hearer believes that p, i.e., $\neg \mathbf{B}_a^{t_2} \mathbf{B}_b^{t_2} p$), it is no longer common ground that p at t_2.

The above time indexed definition of common ground does not apply to all forms of real-life communication. Discourse can stretch out over time so that the speaker's act of producing a particular utterance may be hours, days or years before the hearer's interpretation of the utterance (e.g., in the case of answering machine messages, books, blogposts, letters, etc.). Hence it is possible that the times at which the speaker has propositional attitudes relevant to establishing common ground only partly overlap, or do not overlap at all, with the times at which the hearer has the relevant propositional attitudes. Consider, for instance, my reading of Wittgenstein's *Tractatus*. Since Wittgenstein died long before I was born, whatever he believed, accepted or was committed to act on being true, he did so before I ever believed or accepted anything or was committed to act on anything being true. Hence we never had any attitudes or commitments simultaneously (let alone simultaneous beliefs about what the other currently believes), and thus – on the above common ground definition – there never was any common ground between Wittgenstein and myself.[6]

---

[6]Similar problems concerning delays in communication can pop up in less extreme scenarios. For instance, Halpern and Moses (1990) show that (even if the hearer engages with the speech act while the speaker still has the appropriate attitudes), if there is no fixed time for a message to arrive, there cannot arise common knowledge (e.g., there is no point at which the speaker can be sure that the hearer has received the message and has updated her attitudes. They can be sure if and when they receive a confirmation

A possible strategy is to bite the bullet and maintain that, indeed, there can never exist common ground between a speaker and a hearer if they have no simultaneous attitudes or commitments. However, although it may seem acceptable to maintain that there is no common ground between Wittgenstein and myself (e.g., because I'm not a 'proper' hearer of Wittgenstein's assertions), the same problem of non-simultaneous attitudes can occur on a much shorter time span involving clearly proper hearers. Suppose that a father on his deathbed writes "The Irish crown jewels are hidden in Nouvion" in a letter to his favourite daughter that she is meant to read when he is dead. Intuitively, the statement in the letter is an assertion made by the father with his daughter as proper hearer. However, on the above definition of common ground, it cannot be analysed as such since the speech act cannot have been a proposal to update simultaneous common beliefs between father and daughter with the information that the crown jewels are in Nouvion.[7]

More importantly, the above strict notion of common ground will not do for the purposes of this dissertation. I aim to model our stereotypical engagement with fiction (e.g., the communication between Tolkien and myself through *The Lord of the Rings*) in terms of common ground. We thus need a different, less stringent notion of common ground that can apply to discourse that stretches over time and involves non-simultaneous attitudes or commitments between hearers and (possibly deceased) speakers.[8]

Semanticists that want to apply the notion of common ground to discourse that involves non-simultaneous attitudes or commitments may opt to borrow notions from the logic of distributed asynchronous systems such as ϕ being 'epsilon common knowledge' (Halpern and Moses (1990)) (i.e., everyone will know within ε time units that ϕ, etc.), 'eventual common knowledge' (Halpern and Moses (1990); Fagin et al. (1999)) (i.e., everyone knows ϕ at some point in time, etc.) or 'concurrent common knowledge'

---

message but at this point the sender of that confirmation message cannot be sure that their message has already arrived, etc.).

[7]This scenario also shows that although the issue of non-simultaneous attitudes or commitments arises naturally in cases of deceased unacquainted authors, it is a separate issue from whether speaker and hearer are acquainted.

[8]Some people may be hesitant to apply the term 'common' ground in the above examples precisely because the attitudes are not simultaneous. Since my general aim is to give a uniform analysis of different types of communication in Stalnaker's *common ground* framework, I will stick to this term. An alternative strategy would be to rephrase Stalnaker's theory and define speech acts as proposals to update a '(shared) background', '(shared) ground' or 'context'.

([Panangaden and Taylor](1998)) (i.e., attitude updates do not have to be simultaneous but are causally related in a consistent manner). I leave exploration of what notion of common knowledge would be most useful to future research.

For the purpose of this dissertation I tacitly assume the following simple definition of time indexed common ground that borrows elements from the concept of eventual common knowledge but defines common ground as existing between time slices of conversational participants. p is common ground between speaker a at t_1 and hearer b at t_2 iff:[9]

$$
\frac{\mathbf{B}_a^{t_1} p \quad \mathbf{B}_b^{t_2} p}{\mathbf{B}_b^{t_2} \exists t(t \leq t_2 \wedge \mathbf{B}_a^t p) \quad \mathbf{B}_a^{t_1} \exists t(t \geq t_1 \wedge \mathbf{B}_b^t p)}
$$

$$
\begin{array}{ll}
\mathbf{B}_a^{t_1} \exists t(t \geq t_1 \wedge & \mathbf{B}_b^{t_2} \exists t(t \leq t_2 \wedge \\
\mathbf{B}_b^t \exists t'(t' \leq t \wedge \mathbf{B}_a^{t'} p)) & \mathbf{B}_a^t \exists t'(t' \geq t \wedge \mathbf{B}_b^{t'} p))
\end{array}
$$

$$
\vdots \quad \vdots
$$

Hence some proposition p is not defined as being common ground at a specific point in time or being common ground simpliciter, but rather being common ground between conversational participants at certain points in time. In the case of the father at his deathbed, the father believes some proposition p at t_1; the daughter believes p at t_2; the father believes at t_1 that his daughter will believe that p (at some later time); the daughter believes at t_2 that her father believed that p (at some earlier time); etc. This is enough to say that it is common ground that p between father at t_1 and daughter at t_2.[10] Setting aside issues concerning acquaintance for now, the communication between Wittgenstein (or Tolkien) and myself also falls under this definition; the author had the appropriate beliefs in his time, I have the appropriate beliefs in my time and hence there exists common ground between us.

---

[9]Because this is a definition of common ground between only one speaker and one hearer, it can display an asymmetry in attitudes that is not present in generalized common ground definitions (i.e., between time slices of more than two conversational participants). Assuming that the production of a speech act is always prior to its interpretation, the speaker believes that the hearer *has or will have* certain attitudes, whereas the hearer believes that the speaker *has or had* certain attitudes.

[10]If either conversational participant lacks the appropriate beliefs at some other time (e.g., the daughter no longer believes p at t_3, i.e., $\neg \mathbf{B}_b^{t_3} p$) there exists no common ground between those time slices of the interlocutors (e.g., it is not common ground that p between father at t_1 and daughter at t_3).

2.3.2 Communication without acquaintance

In this subsection I introduce a novel challenge to traditional definitions of common ground concerning unacquainted conversational participants. As shown above, traditional textbook common ground definitions in terms of attitudes seem to involve beliefs about others, i.e., de re beliefs. In other words, they take face-to-face[11] conversations – where conversational participants are acquainted with each other – as a model for communication in general. However, people that are not acquainted can also communicate (e.g., Monk who communicates with his readers). We thus need to further adjust our definition of common ground. To clearly show the worry, I will in this section reformulate the belief-based definition of common ground under a relational analysis of de re attitudes that makes explicit the required acquaintance relations. In the rest of this chapter I will focus on belief-based common ground definitions but the discussed issues and solutions straightforwardly extend to common ground definitions in terms of other cognitive attitudes such as acceptance. It is an open question whether they also extend to a commitment-based conception of common ground.[12]

Relational analysis of de re belief

In the relational analysis of de re belief (Kaplan (1968); Lewis (1979). See also Cresswell and Von Stechow (1982)) if *a* believes de re of *b* that he is *Q*, this means that there is an acquaintance relation between *a* and *b*, and that *a* believes that the person he knows through this acquaintance relation is *Q*. De re beliefs are thus in essence de dicto beliefs paired with an acquaintance relation that links the believer to the res. Such an analysis allows us to account for so-called Ortcutt scenarios (see Quine (1956)): Suppose you and your favourite nephew Nick are driving to your uncle's birthday party in separate cars. Nick cuts you off on the highway but, although you see a glimpse of the driver, you do not realize that it is Nick nor recognize Nick's car. You become angry at the driver and form the de re belief about Nick that

---

[11]Here 'face-to-face' conversations do not require conversational participants to actually be in front of each other. It is sufficient if they know who they are currently talking to (e.g., an online chat conversation with a friend is also a 'face-to-face' conversation).

[12]The main question here is whether we can make a distinction between de re and de dicto commitments; intuitively, '*a* being committed to *b* to act on it being true that *b* is committed to *a* to act on *p* being true' is different from '*a* being committed to *b* to act on it being true that *the hearer* is committed to *a* to act on *p* being true'.

he is a complete ass but lose sight of the car quickly. When entering your uncle's house, however, you are greeted by Nick's enthusiastic embrace and are reinforced in your de re belief about Nick that he is a great guy. Does this mean that you have logically inconsistent de re beliefs, i.e., that Nick is a complete ass and a great guy? The relational analysis of de re beliefs can explain why not: You are acquainted with Nick through the car incident and you believe that the person you saw on the highway is a complete ass. You are simultaneously also acquainted with Nick through regular family gatherings and believe that the person you are acquainted with through family gatherings is a great guy. Hence your de re beliefs – which are comprised of perfectly coherent de dicto beliefs – are not inconsistent.

The relational analysis implies that if a has a de re belief about b, then a has a de se belief (e.g., a has a belief about "the person that *I* saw on the highway"). I follow Lewis (1979) in analysing all attitudes (including de re attitudes) as essentially de se attitudes, i.e., as self-ascription of a property. So if a believes de re of b that he is Q, then a is acquainted with b, and a self-ascribes the property of being such that "the person that I am acquainted with is Q". This is represented as follows:

$$\exists R_1[R_1(a,b) \wedge \mathbf{B}_a^* \lambda i[Q(\imath v[R_1(i,v)])]]$$

In the above formula $\exists R_n[R_n(x,y)]$ means that there is an acquaintance relation R_n such that x is acquainted with y through R_n. $\mathbf{B}_x^* P$ means that x self-ascribes the property P. The term $\lambda x[\phi]$ denotes the property of being an x such that ϕ. I use the letters i, i', i'' etc to denote the de se center, i.e., the self. In the rest of this chapter I will abbreviate $\imath v[R_1(i,v)]$ (i.e., "the person that I am acquainted with through relation R_1") as $\imath R_1^i$.

Here and henceforth I assume that identity is always a salient acquaintance relation; everyone is always acquainted with themselves *as themselves* (de se). In the cases I discuss, identity is the only relevant acquaintance relation we have to ourselves.

Iterative de re beliefs

Iterative de re beliefs further complicate this picture. I adopt Maier's (2009a) analysis of iterative de re attitudes where if a believes de re of b that b believes de re of c that she is Q (i.e., in the earlier notation $\mathbf{B}_a \mathbf{B}_b Qc$), then this entails that a is acquainted with both b and c and that a self-ascribes the property of being such that b is acquainted with c and that the former self-

ascribes the property of being such that "the person that I am acquainted with is Q":

$$\exists R_1[R_1(a,b) \wedge \exists R_2[R_2(a,c) \wedge \mathbf{B}_a^* \lambda i[\exists R_3[R_3(\imath R_1^i, \imath R_2^i) \wedge$$
$$\mathbf{B}_{\imath R_1^i}^* \lambda i'[Q(\imath R_3^{i'})]]]]]]$$

So, if Ann believes de re of Bob that he believes de re of Chrissy that she is cool then [1] Ann is acquainted with both Bob and Chrissy through respectively acquaintance relations R_1 and R_2, [2] Ann believes that the person she is acquainted with through R_1 is acquainted with the person she is acquainted with through R_2 and [3] Ann believes that the person she is acquainted with through R_1 believes that the person he is acquainted with is cool.

We can thus rewrite the common ground definition in terms of de re beliefs as follows (making the required acquaintance relations explicit): p is common ground between speaker a and hearer b iff:[13]

$$
\begin{array}{cc}
\mathbf{B}_a^* \lambda i[p] & \mathbf{B}_b^* \lambda i[p] \\
\hline
\exists R_1[R_1(b,a) \wedge \mathbf{B}_b^* \lambda i[\mathbf{B}_{\imath R_1^i}^* \lambda i'[p]]] & \exists R_1[R_1(a,b) \wedge \mathbf{B}_a^* \lambda i[\mathbf{B}_{\imath R_1^i}^* \lambda i'[p]]] \\
\hline
\exists R_1[R_1(a,b) \wedge & \exists R_1[R_1(b,a) \wedge \\
\mathbf{B}_a^* \lambda i[\exists R_2[R_2(\imath R_1^i, i) \wedge & \mathbf{B}_b^* \lambda i[\exists R_2[R_2(\imath R_1^i, i) \wedge \\
\mathbf{B}_{\imath R_1^i}^* \lambda i'[\mathbf{B}_{R_2^{i'}}^* \lambda i''[p]]]]]] & \mathbf{B}_{\imath R_1^i}^* \lambda i'[\mathbf{B}_{\imath R_2^{i'}}^* \lambda i''[p]]]]]] \\
\end{array}
$$

$$\vdots \quad \vdots$$

So, both speaker and hearer self-ascribe the property of being such that p; both are acquainted with the other and self-ascribe the property of being such that the person they are acquainted with self-ascribes the property of being such that p, etc.

Non-face-to-face conversations

The above reformulation of the traditional definition makes explicit why de re common ground definitions do not apply to non-face-to-face communication. I suggest that we can distinguish (at least) three distinct types of non-face-to-face communication that we do intuitively describe in common ground terminology but that traditional definitions do not apply to. I will

---

[13]Arguably, there is an additional implicit assumption that interlocutors have their iterative de re beliefs through the same constant acquaintance relation.

dub these 'messages in bottles', 'anonymous fan letters' and 'double-blind reviews':[14]

- **Message in a bottle**: conversations where the identity of the speaker is known but that of the hearer isn't and hence (although the hearer has de re beliefs about the speaker) the speaker does not have de re beliefs about the hearer. For instance, suppose you receive and read a message in a bottle from a famous person like Sting. Intuitively, it is common ground between Sting and you that there are monkeys on his desert island. However, here the speaker a is not acquainted with the hearer b (i.e., $\neg\exists R_1[R_1(a,b)]$). Conversations through books, blog posts, broadcasted speeches etc. are cases of such conversations with several hearers.

- **Anonymous fan letter**: conversations where the identity of the hearer is known but that of the speaker isn't and hence (although the speaker has de re beliefs about the hearer) the speaker does not have de re beliefs about the hearer. For instance, intuitively, it is common ground between Sting and the writer of an anonymous fan letter that his voice is like the morning sun. However, here the hearer is not acquainted with the speaker (i.e., $\neg\exists R_1[R_1(b,a)]$).

- **Double-blind review**: conversations where the identity of neither conversational participant is known and neither conversational participant has de re beliefs about the other. For instance, intuitively, it is common ground between reviewer and author in a double-blind peer review process that the submitted paper should not exceed 20 pages. However, in these cases neither speaker nor hearer is acquainted with the other (i.e., $\neg\exists R_1[R_1(a,b)] \wedge \neg\exists R_2[R_2(b,a)]$).

---

[14]Actually, these distinctions raise questions about what constitutes an acquaintance relation. Intuitively, I am acquainted with Monk, but Sting is not acquainted with the writer of the anonymous fan letter because – even though I've never met him – I know Monk through his book, reading about him on Wikipedia, someone referring to him etc. and this is not true for Sting and the anonymous admirer. However, is Sting not in essentially the same way (though maybe in an impoverished sense) also acquainted with the anonymous admirer (or the reviewee with the anonymous reviewer) *through* the love-letter (or the review) (see e.g., Jeshion (2010); Recanati (2009))? Maybe anonymous fan letter conversations or double-blind review conversations (i.e., conversations where (at least) the hearer is not acquainted with the speaker) are in fact not possible. However, such a concession would not dissolve the problem with de re common ground definitions; discourse where the hearer is unknown still seems possible.

In the above types of communication either the speaker is not acquainted with the hearer, the hearer is not acquainted with the speaker, or neither is acquainted with the other. Hence the above iteration of de re beliefs cannot materialize in these cases and so there can exist no common ground between speaker and hearer (i.e., between Sting and the person that finds his message, between Sting and his secret admirer or between anonymous author and reviewer). This is unsatisfactory since these conversations *do* seem to involve producing and interpreting the relevant texts against a shared background. For instance, it is felicitous for the anonymous reviewer to write "Maybe this point relates to Grice's third maxim" but not to write "Maybe this point relates to Bridget's paper" because it *is* common ground what Grice's third maxim is, but *not* who Bridget is.

The same problem arises in a generalized definition of common ground in terms of de re beliefs. If we make all acquaintance relations explicit, such definitions would be rewritten as follows: p is common ground between all conversational participants in some community iff:

$$\forall x (Cx \to \mathbf{B}_x^* \lambda i[p])$$

$$\forall x \forall y ((Cx \wedge Cy) \to (\exists R_1 [R_1(x,y) \wedge \mathbf{B}_x^* \lambda i[\mathbf{B}_{\imath R_1^i}^* \lambda i'[p]]]))$$

$$\forall x \forall y \forall z ((Cx \wedge Cy \wedge Cz) \to (\exists R_1 [R_1(x,y) \wedge \exists R_2 [R_2(x,z) \wedge \\ \mathbf{B}_x^* \lambda i[\exists R_3 [R_3(\imath R_1^i, \imath R_2^i) \wedge \mathbf{B}_{\imath R_1^i}^* \lambda i'[\mathbf{B}_{R_3^{i'}}^* \lambda i''[p]]]]]))$$

$$\vdots$$

In words, everyone in the community self-ascribes the property of being such that p; everyone in the community is acquainted with everyone in the community and self-ascribes the property of being such that the person they are acquainted with self-ascribes the property of being such that p; etc. Here in the cases of two-person conversations through messages in bottles, anonymous fan letters and double-blind reviews, at least one of the conversational participants (the speaker or the hearer) is not acquainted with at least one of the other conversational participants (i.e., $\exists x \exists y (Cx \wedge Cy \wedge \neg \exists R_1 [R_1(x,y)])$) and hence there can be no common ground in the relevant community.

2.4 Redefining common ground

Now that I have spelled out the challenge posed by non-face-to-face conversations to traditional common ground definitions, I will propose two potential fixes to these definitions. The two fixes are inspired by the intuitive idea that, although Monk may not have de re beliefs about (all) his readers individually, he does have general beliefs about 'his readers' and what they believe. In this section I discuss two ways of fleshing out the notion of general belief (i.e., *in sensu composito* and *in sensu diviso*) and propose two definitions of common ground based on these.

2.4.1 Generality *in sensu composito* and *in sensu diviso*

The two notions of common ground that I will propose are inspired by Abelard's distinction between two types of generality (as discussed by Lewis (1969)): *in sensu composito* or 'collective' (Section 2.4.2) and *in sensu diviso* or 'distributive' (Section 2.4.3).[15] If I believe a general rule *in sensu composito* then I have a general *de dicto* belief. For instance, if *a* considers all flowers to be pretty *in sensu composito*, then she believes 'that all flowers are pretty' (i.e., $\mathbf{B}_a \forall x(Fx \rightarrow Px)$ where Fx and Px respectively mean x is a flower and x is pretty). This means that there may be flowers that *a* does not believe to be pretty (for instance because she fails to realize that they are flowers).

Conversely, if I believe a general rule *in sensu diviso* then I have a general disposition to form singular de re beliefs in every relevant situation. For instance, if *a* considers all flowers to be pretty *in sensu diviso*, then she believes of every flower, if she sees it, that it is pretty (i.e., $\forall x(Fx \rightarrow (Sax \rightarrow \mathbf{B}_a Px))$ where Sxy means x sees y)[16]. This means that *a* might not recognize every flower as a flower (might even lack the concept of 'flower' altogether) but still believes of every flower that she comes across that it is pretty.

---

[15]Bermúdez (2003) construes from Braithwaite's (1932) account of generality the following intermediate concept of general belief: if *a* believes that flowers are pretty then *a* believes of every flower that she sees and considers to be a flower, that it is pretty (i.e., $\forall x((Fx \wedge Sax \wedge \mathbf{B}_a Fx) \rightarrow \mathbf{B}_a Px)$). I leave exploration into the merits of extending this notion to a common ground definition for future research.

[16]Contrary to Bermúdez (2003) and Meggle (2003), I represent the fact that *a* has a disposition to form de re beliefs by a conditional: *if a* is in the relevant situation (e.g., sees a flower), then *a* forms the appropriate beliefs.

2.4.2 *In sensu composito* common ground

Definition

First, I will present the *in sensu composito* definition of common ground. An *in sensu composito* understanding of general thought by *a* about the mental states of conversational partners would be as follows: *a* believes (or self-ascribes the property of being such) that 'all conversational partners in the community believe that *p*' (i.e., in our earlier notation: $\mathbf{B}_a \forall x (Cx \rightarrow \mathbf{B}_x p)$, in the present notation: $\mathbf{B}_a^* \lambda i [\forall x (Cx \rightarrow \mathbf{B}_x^* \lambda i'[p])])$. This leads to the following definition of generalized common ground in terms of general de dicto belief: *p* is common ground between all conversational participants in some community iff:

$$\frac{\forall x (Cx \rightarrow \mathbf{B}_x^* \lambda i[p])}{\frac{\forall y (Cy \rightarrow \mathbf{B}_y^* \lambda i[\forall x (Cx \rightarrow \mathbf{B}_x^* \lambda i'[p])])}{\forall z (Cz \rightarrow \mathbf{B}_z^* \lambda i[\forall y (Cy \rightarrow \mathbf{B}_y^* \lambda i'[\forall x (Cx \rightarrow \mathbf{B}_x^* \lambda i''[p])])])}}$$
$$\vdots$$

In words, everyone in the community believes that *p*; everyone in the community believes that everyone in the community believes that *p*; etc. Hence *p* can be common ground in a community even though nobody has any de re beliefs about anyone. All that is required is that people have appropriate beliefs about what 'everyone in the community of conversational participants' believes.

   I speculate that most semanticists that talk about common ground in a large community of interlocutors (e.g., Monk and his readership) tacitly assume (a version of) the *in sensu composito* definition of common ground. In fact, Schiffer (1972), Meggle (2003) and Maier and Semeijn (forthcoming) provide notions of common belief and common ground in a community that are similar to the above *in sensu composito* definition of generalized common ground. However, I assume that problems concerning our inability to form iterative de re attitudes relate to the absence of the relevant acquaintance relations – not to the number of conversational participants per se. Hence, these problems can also occur in a two-person (non-face-to-face) discourse. In other words, the *in sensu composito* definition of common ground is just as relevant when we talk about common ground between one speaker and one hearer, i.e., in a community of two. An *in sensu composito* understanding of general thought about conversational participants would lead to the

following definition of common ground between speaker and hearer in a two-person discourse in terms of de dicto beliefs about 'the speaker' and 'the hearer':[17] p is common ground between speaker a and hearer b iff (where Sx means x is a speaker and Hx means x is a hearer):

$$\frac{\mathbf{B}_a^* \lambda i[p] \quad \mathbf{B}_b^* \lambda i[p]}{\dfrac{\mathbf{B}_b^* \lambda i[\mathbf{B}_{1x[Sx]}^* \lambda i'[p]] \quad \mathbf{B}_a^* \lambda i[\mathbf{B}_{1x[Hx]}^* \lambda i'[p]]}{\mathbf{B}_a^* \lambda i[\mathbf{B}_{1x[Hx]}^* \lambda i'[\mathbf{B}_{1x[Sx]}^* \lambda i''[p]]] \quad \mathbf{B}_b^* \lambda i[\mathbf{B}_{1x[Sx]}^* \lambda i[\mathbf{B}_{1x[Hx]}^* \lambda i''[p]]]}}$$

$$\vdots \quad \vdots$$

So, both speaker and hearer believe that p; the hearer believes that 'the speaker' (whoever it is) believes that p; the speaker believes that 'the hearer' (whoever it is) believes that p; etc. Again, p can be common ground without hearer or speaker forming any de re beliefs. All that is required is that they have the appropriate beliefs about what 'the speaker' or 'the hearer' believes. So – applying the *in sensu composito* notion of common ground to the mini-discourse between Mary and John – when Mary asserts (10) that nobody is coming to the party, she proposes that it becomes the case that she believes that nobody is coming to the party; that John believes this; that she believes that 'the hearer' (whoever it is) believes this; that John believes that 'the speaker' (whoever it is) believes this; etc. Similarly, Mary's speech act (10) is felicitous because Mary believes that there is a party; John believes this; Mary believes that 'the hearer' believes this; John believes that 'the speaker' believes this; etc.

Application to four types of conversation

The *in sensu composito* definition is supposed to be a general definition of common ground that, unlike traditional definitions, applies to all four types of communication distinguished in the previous section (i.e., face-to-face conversations, messages in bottles, anonymous fan letters and double-blind reviews). For reasons of space I only show how this works for the

---

[17]We can derive these iterations from the general common ground definition because, in case speaker a and hearer b are aware that they are in a two-person discourse, a and b are both conversational participants (i.e., Ca and Cb), a believes that 'the hearer' is a conversational participant (i.e., $\mathbf{B}_a^* \lambda i[C1x[Hx]]$), b believes that 'the speaker' is a conversational participant (i.e., $\mathbf{B}_b^* \lambda i[C1x[Sx]]$), etc.

definition of common ground between one speaker and one hearer in a two-person discourse but the derivations for the generalized common ground definitions are similar.

Let's first consider conversations of type 'double-blind review' where neither speaker nor hearer is acquainted with the other. In such cases neither conversational participant has any de re beliefs about the other, nor do they believe the other to have these. I assume that the speaker does have iterative de dicto beliefs about the mental state of 'the hearer' (whoever that may be) and vice versa (e.g., the reviewer believes that 'the hearer' believes that p, the reviewee believes that 'the speaker' believes that p, etc). In other words, the mental states of the conversational participants are properly described by the iteration of de dicto beliefs of the *in sensu composito* definition on p. 30. Hence, on an *in sensu composito* understanding of common ground, it *is* common ground between speaker and hearer that p in this scenario and hence the definition shows how communication can take place.

Next, consider 'anonymous fan letter' conversations, i.e., conversations where the identity of the hearer is known but that of the speaker is not. I assume that also in these types of conversations, both interlocutors have iterative de dicto beliefs about the mental state of 'the hearer' and 'the speaker'. For instance, in the conversation between Sting and his anonymous admirer, Sting believes that 'the speaker' (whoever that may be) believes that Sting's voice is like the morning sun (p) and the admirer believes that 'the hearer' believes that p. Hence in this scenario it is also common ground that p on the *in sensu composito* definition. Apart from the relevant de dicto beliefs, one conversational participant (the speaker) also has de re beliefs about their interlocutor. I assume that in anonymous fan letter conversations, the identity of the hearer is 'known'. This means that both speaker a and hearer b believe de re of b that they are the hearer (i.e., $\mathbf{B}_b^* \lambda i[i = \imath x[Hx]] \wedge \exists R_1[R_1(a,b) \wedge \mathbf{B}_a^* \lambda i[\imath R_1^i = \imath x[Hx]]]$), both believe that the other has this de re belief, etc. Hence although Sting only has de dicto beliefs about the mental state of 'the speaker', the admirer can form de re beliefs about Sting's mental state. Assuming we believe the logical consequences of our beliefs, we can for instance derive that since the admirer believes that 'the hearer' believes that p ($\mathbf{B}_a^* \lambda i[\mathbf{B}_{\imath x[Hx]}^* \lambda i'[p]]$) and the admirer believes de re of Sting that he is 'the hearer', that the admirer also believes de re of Sting (through the same acquaintance relation) that he believes that p: $\exists R_1[R_1(a,b) \wedge \mathbf{B}_a^* \lambda i[\mathbf{B}_{\imath R_1^i}^* \lambda i'[p]]]$). Moreover, Sting and the

admirer both have iterative de re or de dicto beliefs about the others' (de re or de dicto) beliefs about their interlocutor.

'Message in a bottle' conversations (i.e., conversations where the hearer is acquainted with the speaker but not vice versa) are the exact mirror image of anonymous fan letters and hence the same reasoning applies to those. In such cases the hearer has de re beliefs about the speaker but not vice versa.

Last, in face-to-face conversations, both conversational participants are acquainted with the other and have de re beliefs about their interlocutor's mental state. I assume that apart from these de re beliefs, interlocutors also have the appropriate iterative de dicto beliefs about the beliefs of 'the hearer' and 'the speaker'. Hence we can also say that p is *in sensu composito* common ground in these types of conversations.

2.4.3 *In sensu diviso* common ground

Definition

Next, I turn to the *in sensu diviso* version of a common ground definition. To formulate the *in sensu diviso* definition, we need to rewrite in conditional (or *in sensu diviso*) form the relational analysis of de re attitudes. In words, if a believes *in sensu diviso* of b that he is Q, then *if* there is an acquaintance relation from a to b, then a believes (or self-ascribes the property of being such) that the person he knows through this acquaintance relation is Q. This gets translated as $\forall R_1[R_1(a,b) \rightarrow \mathbf{B}_a^* \lambda i[Q \imath R_1^i]]$.

An *in sensu diviso* understanding of a general thought by a about the mental states of conversational partners would be as follows: It is true of all conversational partners in some community that if a is in a relevant situation with the conversational participant, then a believes of this person that they believe that p (i.e., in our earlier notation: $\forall x(Cx \rightarrow (Rax \rightarrow \mathbf{B}_a\mathbf{B}_x p))$ where Rxy means x is in a relevant situation with y). For now, I will assume that 'the relevant situations' are situations where an acquaintance relation obtains[18] so that it is true of all conversational partners in some community that if there is an acquaintance relation from a to the conversational participant, then a believes that the person she is acquainted with believes that p. This gets translated as $\forall x(Cx \rightarrow \forall R_1[R_1(a,x) \rightarrow \mathbf{B}_a^* \lambda i[\mathbf{B}_{\imath R_1^i}^* \lambda i'[p]]])$. This leads

---

[18]We can put further constraints on the relevant types of acquaintance relations. See section 2.5.2.

to the following definition of generalized common ground in terms of conditional de re belief:[19] *p* is common ground between all conversational participants in some community iff:

$$\frac{\forall x(Cx \to \mathbf{B}_x^* \lambda i[p])}{\forall x \forall y((Cx \wedge Cy) \to \forall R_1[R_1(x,y) \to \mathbf{B}_x^* \lambda i[\mathbf{B}_{\imath R_1^i}^* \lambda i'[p]]])}$$

$$\overline{\forall x \forall y \forall z((Cx \wedge Cy \wedge Cz) \to \forall R_1, R_2[(R_1(x,y) \wedge R_2(x,z)) \to}$$
$$\mathbf{B}_x^* \lambda i[\forall R_3[R_3(\imath R_1^i, \imath R_2^i) \to \mathbf{B}_{\imath R_1^i}^* \lambda i'[\mathbf{B}_{\imath R_3^{i'}}^* \lambda i''[p]]]]])$$

$$\vdots$$

So, everyone in the community believes that *p*; for everyone in the community it is true that for everyone in the community, if the one is acquainted with the other, then the one believes that the person they are acquainted with believes that *p*; etc. So, again, no de re beliefs are required for *p* to be common ground in a community. Common ground can exist in a community where no two people are acquainted. It is merely required that people would form the appropriate de re beliefs about one another if they were acquainted (and would believe that the others would as well).

In sensu diviso common ground between speaker *a* and hearer *b* is defined as follows:[20] *p* is common ground between speaker *a* and hearer *b* iff:

$$\frac{\mathbf{B}_a^* \lambda i[p] \quad \mathbf{B}_b^* \lambda i[p]}{\forall R_1[R_1(b,a) \to \mathbf{B}_b^* \lambda i[\mathbf{B}_{\imath R_1^i}^* \lambda i'[p]]] \quad \forall R_1[R_1(a,b) \to \mathbf{B}_a^* \lambda i[\mathbf{B}_{\imath R_1^i}^* \lambda i'[p]]]}$$

$$\overline{\begin{array}{cc} \forall R_1[R_1(a,b) \to & \forall R_1[R_1(b,a) \to \\ \mathbf{B}_a^* \lambda i[\forall R_2[R_2(\imath R_1^i, i) \to & \mathbf{B}_b^* \lambda i[\forall R_2[R_2(\imath R_1^i, i) \to \\ \mathbf{B}_{\imath R_1^i}^* \lambda i'[\mathbf{B}_{\imath R_2^{i'}}^* \lambda i''[p]]]]] & \mathbf{B}_{\imath R_1^i}^* \lambda i'[\mathbf{B}_{\imath R_2^{i'}}^* \lambda i''[p]]]]] \end{array}}$$

$$\vdots \quad \vdots$$

So, both speaker and hearer believe that *p*; if the hearer is acquainted with the speaker, then the hearer believes that the person they are acquainted

---

[19] Meggle (2003) provides an alternative '*in sensu diviso*' version of a generalized common ground definition that boils down to the traditional definition in terms of de re beliefs. This version does not seem to do justice to the fact that a general *in sensu diviso* belief involves a *disposition* to form de re beliefs. Moreover, it suffers from the problems described in section 2.3.2 and hence is, in the context of this discussion, not the most interesting version of an *in sensu diviso* definition.

[20] We arrive at these iterations from the general common ground definition because *a* and *b* are conversational participants (i.e., *Ca* and *Cb*).

with believes that *p*; if the speaker is acquainted with the hearer, then the speaker believes that the person they are acquainted with believes that *p*; etc. So *p* can be common ground between speaker and hearer even though neither has any de re beliefs about the other. All that is required is that they would form the right de re beliefs about each other if they were acquainted (and would believe the other would do so as well).

So when Mary asserts (10) that nobody is coming to the party, she proposes that it becomes the case that she believes that nobody is coming to the party; that John believes this; that, if she is acquainted with John, then she believes of John that he believes this; that, if John is acquainted with Mary, he believes of Mary that she believes this; etc. Similarly, Mary's speech act (10) is felicitous because Mary believes that there is a party; John believes this; if Mary is acquainted with John, then she believes that of John that he believes this; if John is acquainted with Mary, then he believes of Mary that she believes this; etc.

A potential worry with the above *in sensu diviso* definitions is that they universally quantify over acquaintance relations, making the formulas trivially true when there are no relevant acquaintance relations. On the standard material implication analysis of conditionals, quantification over an empty domain is trivially true (e.g., given that there are no unicorns, it is true of all things that if they are a unicorn, then they are pink, black, robots, etc.). For the generalized de re and *in sensu composito* definitions (which also involve quantification) this does not lead to serious difficulties (i.e., it implies that if there are no conversational participants (i.e., $\neg\exists x Cx$), then everything is trivially common ground). However, consider the *in sensu diviso* definition for common ground between speaker *a* and hearer *b*. If *a* is not acquainted with *b* (i.e., $\neg\exists R_1[R_1(a,b)]$), then it is *trivially* true that for all acquaintance relations, if they are from *a* to *b*, then *a* believes de re of *b* that they believe that *p* (i.e., $\forall R_1[R_1(a,b) \rightarrow \mathbf{B}_a^*\lambda i[\mathbf{B}_{1R_1^i}^*\lambda i'[p]]]$). The definition thus implies that if *a* and *b* are not acquainted, everything is trivially common ground between them. This is problematic because the *in sensu diviso* definition is supposed to apply to cases of non-face-to-face conversations and tell us whether some proposition is common ground in those situations or not.

To overcome this issue with the *in sensu composito* common ground definitions, we need a non truth-functional account of conditionals: Given that the antecedent (*p*) is false, the conditional ($p \rightarrow q$) can still be true or false. For instance, we can take '$\rightarrow$' (at least when embedded under a quantification

over acquaintance relations) to denote an 'indicative conditional' and adopt the Stalnaker/Lewis (Stalnaker (1968); Lewis (1973)) possible world analysis: $p \rightarrow q$ is true iff in the closest (or most similar) possible worlds to the actual world where p is true, q is also true.[21] If p is in fact true, the actual world is the closest possible p-world to the actual world (and hence q must be true in order for the conditional to be true). Given this analysis, even if a is not acquainted with b, then still $\forall R_1[R_1(b,a) \rightarrow \mathbf{B}_b^* \lambda i[\mathbf{B}_{iR_1^i}^* \lambda i'[p]]]$ may be true or false: True if the closest possible worlds where a *is* acquainted with b are such that a believes de re of b that they believe that p. False if this isn't the case.

Application to four types of conversations

Again, I will show how the four distinguished types of face-to-face and non-face-to-face communication are supposed to fit into the *in sensu diviso* definition.

First, I will consider non-face-to-face conversations where neither speaker nor hearer is acquainted with the other as in double-blind peer reviewing. In such conversations, neither interlocutor has any de re beliefs about the other, nor believes the other to have these.[22] However, I assume that the interlocutors do have a disposition to form the appropriate de re beliefs; if they had been acquainted, they would have formed the relevant iteration of de re beliefs about each other (e.g., reviewer a would have believed de re of the author b that they believe that the submitted paper should not exceed 20 pages (p): $\forall R_1[R_1(a,b) \rightarrow \mathbf{B}_a^* \lambda i[\mathbf{B}_{iR_1^i}^* \lambda i'[p]]]$).[23] In other words, the mental states and dispositions of interlocutors in conversations of this type satisfy the iteration of mental states and dispositions described in the

---

[21]See chapter 4 for Lewis' application of this analysis of conditionals to counterfactuals and in particular the fiction operator 'In f, ϕ'.

[22]They may, as described above, have de dicto beliefs about the mental state of 'the speaker' or 'the hearer'.

[23]Arguably this may not be true for *all* possible acquaintance relations. For instance, if a saw b on the beach but does not realize that that was the person they are reviewing, a probably will not believe de re of b that they believe that the submitted paper should not exceed 20 pages. This suggests that an *in sensu diviso* definition in fact requires further constraints so that it only quantifies over acquaintance relations 'in the context of the conversation'. See section 2.5.2 for further discussion.

in sensu diviso definition on page 33. Hence *p* is common ground in this type of conversation under the *in sensu diviso* definition.

Next, I will consider conversations of type 'anonymous fan letter' where the speaker is acquainted with the hearer and has de re beliefs about them but not vice versa. I assume that in such a situation, although the hearer does not have any de re beliefs about the speaker, they do have a disposition to form these, had they been acquainted. The speaker, on the other hand, has de re beliefs about the hearer and their mental state (e.g., the anonymous admirer *a* believes de re of Sting *b* that he believes that his voice is like the morning sun (*p*): $\exists R_1[R_1(a,b) \wedge \mathbf{B}_a^* \lambda i[\mathbf{B}_{1R_1^i}^* \lambda i'[p]]])$ However, from the fact that there is such an acquaintance relation we cannot derive that *a* has the appropriate beliefs for *all* acquaintance relations. Hence, to make his definition work, the defender of the *in sensu diviso* definition will have to assume that in 'anonymous fan letter' conversations, there is not only this acquaintance relation from *a* to *b*, but *a* also still has a general disposition (that the existent acquaintance relation and accompanying beliefs are in line with) to form the appropriate beliefs if he were otherwise acquainted with *b*: $\forall R_1[R_1(a,b) \rightarrow \mathbf{B}_a^* \lambda i[\mathbf{B}_{1R_1^i}^* \lambda i'[p]]]$.[24] Hence on the *in sensu diviso* definition *p* is common ground in conversations of this type.

'Message in a bottle' conversations are the exact mirror image of 'anonymous fan letter' conversations (where the hearer, rather than the speaker, has de re beliefs about their interlocutor) and work similarly.

Last, in face-to-face conversations, both interlocutors have iterated de re beliefs about each other. Again, we assume that in addition to having de re beliefs about their interlocutor, speaker and hearer also have a general disposition to form the appropriate de re beliefs if they were otherwise acquainted. Hence the *in sensu diviso* definition also applies to these kinds of conversations.

---

[24]An alternative strategy for the defender of an *in sensu diviso* definition is to assume a weaker version of the definition which requires *a* to either be acquainted and have the appropriate beliefs, or have a disposition to have these beliefs if *a* were acquainted:$\exists R_1[R_1(a,b) \wedge \mathbf{B}_a^* \lambda i[\mathbf{B}_{1R_1^i}^* \lambda i'[p]]] \vee \forall R_1[R_1(a,b) \rightarrow \mathbf{B}_a^* \lambda i[\mathbf{B}_{1R_1^i}^* \lambda i'[p]]]$.

2.5 Comparison

Now that I have presented two possible strategies to improve upon traditional definitions of common ground, we can try to decide between them. Most importantly, we can compare how well they fit our intuitive understanding of common ground. Lewis (1969) has argued that a general rule is convention only if people believe it *in sensu diviso*, i.e., only if people respond in accordance with the rule in all relevant instances (whether they are aware of this or not). Similarly, one could argue that *q* is only truly common ground when people form the appropriate de re beliefs about conversational participants whenever they are acquainted with them. However, it is not obvious that Lewis' reasoning extends to the concept of common ground. I will first discuss what implications the different definitions have for the notion of felicity. Then I will discuss the case of the shy acquaintance that may help us decide between the two definitions.

2.5.1 Felicity and counterfactual success

Remember that common ground is not just what we propose to update with assertions, but also determines what speech acts are felicitous or appropriate in a particular context. For instance, reconsider the mini-discourse between Mary and John. If it is common ground that there is a party (*q*), then Mary can felicitously assert (10), i.e., that nobody is coming to the party (*p*). Otherwise, her speech act would not be not felicitous. The *in sensu diviso* definition is crucially different from traditional common ground definitions in terms of iterative de re beliefs and *in sensu composito* common ground definitions because it defines context (and hence felicity) in terms of *conditional* beliefs, rather than actual beliefs. Hence, whereas for the de re and *in sensu composito* definitions the felicity of a speech act *p* in a context *c* is dependent on whether *p* will lead to successful communication in *c*, for the *in sensu diviso* definition it is dependent on whether *p* leads to successful communication in the context of the face-to-face version of the relevant conversation *c'*, which may or may not be identical to *c*.

To illustrate, consider Mary and John's discourse under the traditional de re and *in sensu composito* definitions of common ground. Suppose that it is common ground that there is a party (*q*). On the de re common ground definition, if *q* is common ground, then Mary and John are in a face-to-face conversation and Mary believes de re of John that he believes that *q*. On the

in sensu composito definition, if *q* is common ground, then Mary has (possibly in addition to de re beliefs about John) the de dicto belief that 'the hearer' believes that *q* in all four distinguished types of conversation. Hence, if Mary wants to engage in cooperative communication, she would in fact say that nobody is coming to the party (*p*) (which presupposes *q*) because she would expect her interlocutor to understand her utterance of *p*. Similarly, if *q* is common ground, John would be able to interpret Mary's assertion in face-to-face conversations and (under the *in sensu composito* definition) in all distinguished types of non-face-to-face conversations (because he either believes de re of Mary that she believes *q* or because he believes that 'the speaker' believes that *q*). If *q* is not common ground, either Mary will not say *p* (e.g., because she doesn't believe that her interlocutor believes that *q*) or John will not be able to interpret *p* (e.g., because he does not believe his interlocutor believes that *q*). Hence on both definitions, if the speech act *p* is predicted to be felicitous (because it is common ground that *q*), this implies that *p* will also *actually* lead to successful communication.

The *in sensu diviso* definition is crucially different because it provides a notion of felicity that is tied to *conditional* success. For instance, in a non-face-to-face conversation between Mary and John of type 'double-blind review' (where neither is acquainted with the other) it can, in theory, be common ground that there is a party (*q*) even though Mary believes that 'the hearer' does not believe that *q*, John believes that 'the speaker' does not believe that *q*, etc. Still it would be common ground that *q* as long as it is true that *if* Mary and John were acquainted, they would form the right de re beliefs about one another. Mary's assertion that nobody is coming to the party (*p*) is thus predicted to be felicitous even though *p* would not actually lead to successful communication; Mary would not actually utter *p* if she were trying to engage in cooperative communication (because – although if she were acquainted with John, she would believe of him that he believes that *q* – she does not currently believe that 'the hearer' believes that *q*) and, even if Mary did utter *p*, John would not even be able to interpret *p* (because – even though if he were acquainted with Mary, he would believe of her that she believes that *q* – he does not currently believe that 'the speaker' believes that *q*).

In other words, it seems that the intuitive idea behind the *in sensu diviso* definition is to take face-to-face conversations as a model for communication and define felicity in terms of what speech acts would lead to successful communication in the face-to-face version of the relevant conversation. Hence

we end up with an essentially different notion of felicity on the *in sensu diviso* definition than on de re and *in sensu composito* definitions. Whether some speech act is felicitous in some context thus becomes a separate issue from whether the speech act will actually lead to successful communication in that context. Some people may find this an unintuitive divorce and hence may want to avoid such a notion of common ground. Assuming we want a definition of common ground that also applies to non-face-to-face conversations, they are thus pushed into the direction of the *in sensu composito* definition. This is the notion of common ground that I will assume in subsequent chapters. I will end this chapter with a discussion of a particular kind of conversation that can help us tease apart the two available notions of common ground for non-face-to-face conversations (*in sensu composito* and *in sensu diviso*) and hence help us decide which fits our intuitive understanding better.

2.5.2 The shy acquaintance

To clearly see the difference between the two notions of common ground (and hence the two notions of the felicity of speech acts), suppose Thea signed up for a program where participants send weekly 'diary entries' to an anonymous reader. Unbeknownst to Thea, she has been sending her diary entries to her favourite nephew Nick who signed up to the program to be an anonymous reader. Nick quickly found out that he is receiving his aunt's diary entries but has so far been too shy to admit this to Thea. Suppose Thea is interested in architecture and her latest diary entry contained the information that *The Sims* was originally designed as an architecture simulator (q). According to the *in sensu diviso* definition of common ground, it is *not* common ground between Thea t and her nephew n that q – Thea is acquainted with the hearer but does not believe of him (under that acquaintance relation) that he believes that q (i.e., $\exists R_1[R_1(t,n) \land \neg\mathbf{B}^*_t \lambda i[\mathbf{B}^*_{\imath R^i_1} \lambda i'[q]]]$).

According to the *in sensu composito* definition, it *is* common ground between Thea and her nephew that q – Thea has the appropriate beliefs about what 'the hearer' believes (i.e., $\mathbf{B}^*_t \lambda i[\mathbf{B}^*_{\imath x[Hx]} \lambda i'[q]]$) and the hearer is her nephew Nick (even though Thea does not recognize him as such).

An anonymous Amsterdam Colloquium 2019 reviewer has judged a comparable case to form an argument in favour of the *in sensu diviso* definition: If Thea would meet her nephew at her husband's birthday party, surely

she would not be licensed to base the production of her speech acts on the assumption that q is shared background knowledge. For instance, it would not be correct for her to say (11):

(11)   I gave my husband that architecture simulation game.

since she does not expect her nephew to understand what game she is talking about. Similarly, the linguistic behaviour of the nephew will be as if he has no idea what game Thea is talking about (since he does not want to expose himself as a reader).

I agree with the reviewer's judgement but argue that in the above example conversation it would also not be common ground that q on an *in sensu composito* understanding. This is because during the birthday conversation Thea does not believe that 'the hearer' (of that conversation)[25] believes that q. The difference between the *in sensu composito* and the *in sensu diviso* definition in fact only comes out in conversations where Thea believes that 'the hearer' (of that conversation) believes that q. So the relevant situation to consider is one where Thea is writing another diary entry; here Thea believes that 'the hearer' believes that q (because the previous entry contained this information) and she is acquainted with the hearer (i.e., her nephew) but does not believe of him that he believes that q. Although intuitions may vary concerning this and related cases, I take these considerations to form a prima facie argument for the *in sensu composito* definition of common ground, i.e., in this situation it *is* common ground between Thea and her nephew that q. For instance, it would be felicitous for Thea to write (11) in such a situation precisely because she expects 'the reader' to understand what she is talking about (and the reader has the appropriate beliefs about 'the writer' as well).

A possible response to the above argument in favour of the *in sensu composito* definition is to add further constraints on acquaintance relations in the *in sensu diviso* definition. Intuitively, the mere existence of an acquaintance relation is not enough to establish that conversational participants are in a 'relevant situation' for common ground. Reconsider a who believes flowers to be pretty *in sensu diviso*. For a to be in a relevant situation for establishing

---

[25]This indexical element implicit in the *in sensu composito* definition also explains how for instance two people can have multiple common grounds, i.e., they may not realize that their interlocutor in one communicative exchange is the same as their interlocutor in some other communicative exchange. Common ground is thus in fact defined as something that exists not just between interlocutors but between interlocutors in a particular communicative exchange.

general *in sensu diviso* belief, it is not enough for *a* just to be acquainted with a flower. Arguably, if I hit *a* on the back of the head with a bouquet of flowers, she is acquainted with the flowers but not in a relevant situation because she does not see the relevant flowers. In other words, the fact that *a* does not consider those flowers to be pretty does not disprove her general *in sensu diviso* belief that flowers are pretty. Similar constraints could be put on the relevant acquaintance relations necessary for establishing *in sensu diviso* common ground. Although it is challenging to make such notions precise, we could develop a version of *in sensu diviso* common ground that requires that conversational participants are acquainted *in the context of the relevant conversation*. In other words, we don't just consider counterfactual situations where conversational participants are acquainted simpliciter but rather where the relevant conversation is face-to-face.[26] Such an account would arguably predict that Thea's speech act *is* felicitous because Thea is not acquainted with her nephew in the context of writing the diary entries and would believe de re of him that he believes that *q* if she were. I leave exploration of how to make the notion of 'being acquainted in the context of the relevant conversation' more precise and the differences between this adjusted *in sensu diviso* definition and the *in sensu composito* definition to future research.

2.6 Conclusions

The main contribution of this chapter is a proposal of two potential refinements of traditional definitions of common ground in terms of de re attitudes in order to make them applicable to non-face-to-face communication. These come forth from Abelard's distinction between generality *in sensu composito* (i.e., speaker and hearer have iterative de dicto beliefs about the mental states of 'the speaker' and 'the hearer') and *in sensu diviso* (i.e., speaker and hearer form iterative de re beliefs about the mental states of speaker and hearer if they are acquainted with them). I have shown that the *in sensu diviso* definition makes the felicity of speech acts depend on

---

[26]Similarly, the four different types of discourse that have been distinguished in section 2.3.2 can be reformulated with these further constraints. For instance, a conversation is of the type 'double-blind review' iff conversational participants are not acquainted *in the context of the conversation*. Hence close colleagues may engage in such a non-face-to-face conversation even though they are otherwise acquainted.

counterfactual successful communication, i.e., what speech acts would lead to successful communication in the face-to-face version of the relevant conversation. Therefore this definition is less suitable for my purposes. I will henceforth in this dissertation assume the *in sensu composito* definition of common ground. Lastly, I have argued that the case of the shy acquaintance forms a prima facie argument in favour of an *in sensu composito* definition.

3 Unofficial common grounds

This chapter is a rewritten and significantly expanded version of the first three sections of 'A Stalnakerian analysis of metafictive statements' in *Proceedings of 21ˢᵗ Amsterdam Colloquium*. A small part of subsection 3.4.2 is adapted from 'Revisiting the 'wrong kind of object' problem' which is a co-authored paper with Prof. Dr. Edward N. Zalta in *Organon F*. The most substantial differences between this chapter and the proceedings paper include: First, the inclusion of an introduction to the basic puzzle of fictional discourse as the need to quarantine fictional content (section 3.1). Second, the addition of a formalisation of the unofficial common ground accounts in the DRT formalism and an introduction of DRT (section 3.3.2). Third, the inclusion of a more elaborate discussion of parafictional discourse and different semantic analyses of it.

3.1 Introduction

The current and the next chapter deal with the central puzzle that has kick-started the study of the semantics of fiction: How do we model the difference between fictional and non-fictional talk? In a dynamic semantic framework this boils down to the question of what kind of context updates result from assertions on the one hand and so-called 'fictional statements' on the other. Reconsider the example of Tolkien's assertion about his friend C.S. Lewis (1):

(1)   C.S. Lewis was born in Belfast.

Such non-fictional speech acts have to be distinguished from fictional statements, i.e., statements that are part of a fictional narrative. For instance, Tolkien's written statement (2) which is a quote from *The Hobbit*:

(2)   In a hole in the ground there lived a hobbit.

Whereas assertions such as (1) are statements about the real world that are actually true or false, fictional statements are usually analysed as neither true nor false. Rather, they are fictional truth-makers; Tolkien's writing of (2) makes it true in the world of *The Lord of the Rings* that a hobbit lived in a hole in the ground.

As we have seen in the previous chapter, in Stalnaker's dynamic framework, assertions are defined as proposals to update the common ground between speaker and hearer. For instance, by asserting (1), Tolkien proposes that it becomes common ground that C.S. Lewis was born in Belfast. Because the common ground framework is modelled after cooperative information exchanges – where people share beliefs through assertions to increase their shared background – it is challenging to model conversations in which people say things they believe to be (strictly) false, such as when people tell a fictional story (e.g., Tolkien does not actually believe that a hobbit lived in a hole in the ground). To illustrate, fictional statements such as (2) cannot be modelled as assertions, i.e., as simple updates of the common ground. On a simple de re belief-based conception of common ground[1] this would entail that, after fictional statement (2), it is common belief between Tolkien a and myself b that in a hole in the ground there lived a hobbit (p), i.e., $\mathbf{B}_a p$, $\mathbf{B}_b p$, $\mathbf{B}_a \mathbf{B}_b p$, $\mathbf{B}_b \mathbf{B}_a p$, etc. Such an analysis quickly runs into difficulties.

First of all, p leads to an inconsistent common ground. Arguably, p implies that hobbits exist (q) and hence this also becomes common ground (i.e., $\mathbf{B}_a q$, $\mathbf{B}_b q$, etc.). However, before engaging with *The Lord of the Rings* it was probably already common ground between Tolkien and myself that there are no such creatures as hobbits (i.e., $\mathbf{B}_a \neg q$, $\mathbf{B}_b \neg q$, etc.). An update that leads to an inconsistent common ground is, in itself, not necessarily problematic if it is clear how to resolve the inconsistency (see section 3.3.1 and chapter 8). However, if we model fictional statements as assertions, inconsistencies *are* problematic because it becomes unclear how to proceed; it is unclear which of the inconsistent propositions should be removed from our common ground to make it consistent again. While it is very unintuitive to remove q from the common ground because it is so obviously true to Tolkien and myself that hobbits do not exist, removing p from the common ground defeats our original purpose of modelling the updates resulting from fictional statements.

---

[1] The same issues that are described below – concerning inconsistency and quarantining – come up on *in sensu composito* or acceptance-based conceptions of common ground.

Second, even if the proposition expressed by some fictional statement is not inconsistent with our previous common ground in any obvious way (e.g., some of the propositions that are expressed by a 'realistic' fictional narrative such as Austen's *Pride and Prejudice* 'might as well have been true'), we still would not want to admit that this proposition *truly* becomes common ground. At least, this fictional information is not common ground in the same sense as it is common ground between Tolkien and myself that C.S. Lewis was born in Belfast. We do not mix fact and fiction in this way. Rather, fictional content should somehow be 'quarantined' from non-fictional content.

In the current chapter I will discuss and present formalisations of existing Stalnakerian accounts that model fictional discourse and the quarantining of fictional content: Eckardt's (2014) and Stokke's (2013, 2018) 'unofficial common ground accounts' (section 3.2). I will critique both accounts on how they deal with two conflicting intuitions concerning fictional content. First, that fictional truths (e.g., that hobbits exist) are only accepted while we engage with the fiction. Second, that we do, somehow, retain information about fictional truths even after engaging with the fiction. This allows us to properly interpret a continuation of the fictional discourse after a break and engage in so-called 'parafictional discourse' (section 3.4).

3.2 Unofficial common grounds

Eckardt's (2014) and Stokke's (2013, 2018) unofficial common ground accounts are Stalnakerian accounts that provide a way to separate fictional content from non-fictional content.[2] In both Eckardt's linguistically motivated approach and Stokke's philosophically motivated approach, a sharp distinction is drawn between fiction interpretation and non-fiction interpretation. This is in line with the traditional and dominant theory of fiction (e.g., Currie (1990); Walton (1990)) according to which fiction interpretation (which involves the cognitive attitude of imagination) is fundamentally different from non-fiction interpretation (which involves belief). Whereas an assertion that expresses a proposition *p* is a mandate to *believe p*, a fic-

---

[2]See Keiser (forthcoming) for a recent critical discussion of Stokke's account of assertion and the distinction between official and unofficial common grounds. See also Green (2017) who also suggested that modeling fictional discourse requires positing multiple common grounds between interlocutors.

tional statement that expresses *p* is a mandate to *imagine p* in the game of make-believe licensed by the fictional narrative.

Likewise, in Stokke's and Eckardt's Stalnakerian frameworks, assertions and fictional statements update the common ground in fundamentally different ways. Assertions are defined as proposals to update, in Stokke's terminology, the 'official common ground'. This is the set of mutually presupposed propositions concerning actual states of affairs. For instance, after assertion (1) the *official* common ground between Tolkien and his addressee is updated with the proposition that C.S. Lewis was born in Belfast. Fictional statements, on the other hand, are defined as proposals to update or create an 'unofficial common ground'[3] related to the relevant fictional narrative. This is the set of propositions that are mutually presupposed by the addressee and author of a fictional story while engaging with the fiction, i.e., those that are updated with the propositions expressed by the narrative. For instance, after fictional statement (2), the *unofficial* common ground between Tolkien and his reader that is specifically related to Tolkien's *The Lord of the Rings* saga is updated with the proposition that there lived a hobbit in a hole in the ground. Because we normally engage in different fictional narratives, a typical 'complete common ground' between two people will contain one official common ground concerning actual states of affairs, and several unofficial common grounds related to different fictions (e.g., the common ground between a friend and myself may consist of one official common ground, a *The Lord of the Rings* unofficial common ground, a *Harry Potter* unofficial common ground, a *Pride and Prejudice* unofficial common ground, etc.).[4]

Unofficial common grounds have to be construed as acceptance-based; Tolkien and I do not actually *believe* but merely *accept* that there is a hobbit called Frodo, believe that the other accepts this, etc. The separation of unofficial common grounds and the official common ground allows us to construe the official common ground (for now)[5] as belief-based; Tolkien

---

[3]In Eckardt's terminology: proposals to update or create STORY$_0$.

[4]Alternatively, following the 'fragmented mind' programme (see e.g., David (2015)), one could formulate an account involving *one* compartmentalized common ground. What are unofficial common grounds in Stokke's and Eckardt's accounts, are different compartments related to different fictions in this framework. Beliefs concerning actual states of affairs (part of the official common ground in the unofficial common ground accounts) are also structured in compartments of the same common ground.

[5]See chapter 5.

and I both *believe* that C.S. Lewis was born in Belfast, believe that the other believes this, etc.

3.3 Formalisation

3.3.1 Sets of propositions

We can represent the complete common ground (C) as a *n*-tuple of one official common ground (C_0), and several numbered unofficial common grounds ($C_1, ..., C_n$). Here both the official common ground and unofficial common grounds are defined as sets of possible world propositions (rather than sets of possible worlds):

$$C = \langle C_0, C_1, ..., C_n \rangle$$

Assertions are defined as proposals to update ($*$) the official common ground:

$$C +_A p = \langle C_0 * p, C_1, ..., C_n \rangle$$

To model how fictional statements update unofficial common grounds, we must distinguish between two cases: Either a fictional statement is a proposal to update an already existing unofficial common ground (e.g., when continuing to read *The Lord of the Rings*), or a fictional statement is a proposal to create a new unofficial common ground (C_{BASE}) and update this common ground (e.g., when starting to read a new fictional novel):

$$C +_{F^i} p = \begin{cases} \langle C_0, C_1, ..., C_{i-1}, C_i * p, C_{i+1}, ..., C_n \rangle & \text{if } 1 \leq i \leq n, \\ \langle C_0, C_1, ..., C_n, C_{BASE} * p \rangle, & \text{otherwise.} \end{cases}$$

This formalisation raises two questions: First, what exactly is the content of C_{BASE}? In other words, what is mutually presupposed by the addressee and author of a fictional story when starting to engage in a new fictional narrative? Is C_{BASE} a copy of the official common ground ($C_{BASE} = C_0$), a tabula rasa ($C_{BASE} = \varnothing$), or something in between (see e.g., Lewis (1978), Ryan (1980) or Lamarque (1990))? In the formalisations presented here I assume that C_{BASE} is a copy of the official common ground between speaker and hearer and hence contains all mutually presupposed propositions concerning actual states of affairs. Assuming that C_{BASE} is a tabula rasa, or

something in between a tabula rasa and a copy of the official common ground (cf. Eckardt (2014), see also footnote 7), is also compatible with the unofficial common ground accounts but would lead to different formalisations.

Assuming that C_{BASE} is a copy of the official common ground enables us to resolve anaphoric links in the unofficial common ground, so when a novel mentions terrorist attacks in Paris, we already have a 'discourse referent' for that city, and all kinds of background information predicated thereof. This is in line with the idea that we're never interpreting a text in a vacuum, but understand it against a background or importation of factual information about the actual world, as in Lewis' (1978) counterfactual analyses of truth in fiction, Ryan's (1980) Principle of Minimal Departure, Walton's (1990) Reality and Mutual Belief Principles, and especially Friend's (Friend (2017)) Reality Assumption. In all these theories, fictional worlds are assumed to be as much as possible like (or, in Lewisian terms 'as close as possible to') the real world as the story permits.[6] Assuming that C_{BASE} is a copy of the official common ground not only allows us to import information into unofficial common grounds that we intuitively find true in for instance *The Lord of the Rings* (e.g., that the sun rises in the east) but also information that we may find difficult to accept as true in *The Lord of the Rings* (e.g., that Paris is the capital of France or that Tolkien wrote a book called *The Lord of the Rings*). However, the alternatives (i.e., analysing C_{BASE} as a *tabula rasa* or something in between) are very difficult to make precise. Both options have to involve some kind of accommodation of the information necessary to appropriately interpret the fictional narrative.[7] The difficulty lies in specifying exactly what information should enter the common ground and what information shouldn't. For example, in order to appropriately interpret the fictional story that mentions terrorist attacks in Paris, the unofficial common ground will at least have to include the information that Paris is the capital of France. But this information is only comprehensible if it also contains background information about what France is. This background information will then inevitably refer to yet many other things that again need further accom-

---

[6] See Franzén (forthcoming) for an in depth discussion of the Reality Principle.

[7] In fact, Eckardt opts for an analysis where new unupdated unofficial common grounds are something in between a copy of the official common ground and a tabula rasa; they specify a very unrestricted set of worlds $STORY_0$ (e.g., it is not common ground that frogs cannot speak at the start of a fictional discourse) that is updated with appropriate content as the fictional discourse proceeds.

modation of information. Hence it is unclear where such accommodation processes end.

The second question raised by the formalism is about the $*$ operator. Updating a common ground with some proposition p is usually formalised as $C \cup \{p\}$ rather than $C * p$, when common grounds are defined as sets of propositions ($C \cap p$ when common grounds are defined as sets of possible worlds). However, especially with fictional statements, simply joining sets of propositions can lead to inconsistent common grounds. For instance, given that we assume C_{BASE} to be a copy of the official common ground, C_{BASE} will contain the information that hobbits do not exist.[8] Thus when we come across (2) when we start to read *The Hobbit*, taking the union of the sets of propositions leads to an inconsistent unofficial common ground. We require an operator that is suitable for such inconsistent updates and tells us how to resolve them (e.g., a belief revision operator, a Lewisian operator or a probability distribution update). Such an operator would in the above case of reading *The Hobbit* have to entail that after an update with (2), it *is* unofficial common ground that there lived a hobbit in a hole in the ground and *not* unofficial common ground that hobbits do not exist. I will elaborate on one such belief revision update mechanism developed by Maier and Semeijn (forthcoming) for modelling such fiction updates in chapter 8. For now we may simply assume that $*$ denotes the kind of operator described.

3.3.2 Discourse representation structures

In most of this dissertation I will use the box notation of Discourse Representation Theory (DRT) developed by Kamp (1981) to represent common grounds. In this subsection I will first briefly introduce the basics of the DRT formalism[9] and then use it to represent the unofficial common ground accounts.

---

[8]In case the reader finds it unintuitive that this is common ground (and hence that Tolkien also accepted this) at the start of the fictional discourse, see the discussion in section 8.2.

[9]For a more elaborate introduction to basic DRT syntax and semantics see Geurts et al. (2016); Kamp and Reyle (1993).

Introducing DRT

DRT offers a formalism to model dynamic context updates of (multi-sentence) discourse.[10] This context is represented as a structured entity in so-called Discourse Representation Structures (DRS's). Mirroring the debate in dynamic semantics over what a conversational context exactly is, DRS's have been used to represent the context interpreted as a Stalnakerian common ground (e.g., Heim (1982); Groenendijk and Stokhof (1991); van der Sandt (1992)) and have been used to represent the context interpreted as an agent's individual mental state (e.g., Geurts (1999); Kamp (2015); Maier (2017)). As the present aim is to define speech acts in terms of common ground updates, here DRS's represent Stalnakerian common grounds.[11]

To illustrate DRT's box notation, consider assertion (1) in a straightforward cooperative information exchange. In DRT, noun phrases or 'NP's' (e.g., the proper name 'C.S. Lewis') in a discourse are mapped to 'discourse referents' placed under several conditions. A common ground updated with the proposition that C.S. Lewis was born in Belfast is represented as follows:

(12)

x
Lewis(x)
born.in.Belfast(x)

The top part of the DRS introduces a discourse referent. We may think of this as akin to existential quantification. Box (12) thus contains the information that there is an x. The bottom part of the DRS introduces conditions that specify properties of and relations between discourse referents. Hence, (12) also contains the information that x is named C.S. Lewis and that x was born in Belfast.

Subsequent assertions will update DRS (12) by adding more discourse referents and conditions as the discourse unfolds. For instance, suppose Tolkien continued his discourse with (13):

(13)   He wrote *The Chronicles of Narnia*.

---

[10]A similar theory has been developed by Heim (1982) independently.

[11]Later in this dissertation, when I focus on situations where the hearer's and the speaker's conceptions of what is common belief diverge (e.g., successful deceptive lies (chapter 5) or cases of unreliable narration (chapter 8)), I will suggest that it is more interesting to model part of the hearer's beliefs, i.e., what they consider to be common ground.

The DRT formalism allows us (amongst other things) to model how and when anaphoric references in an ongoing discourse work. First, we update the DRS with the information expressed in (13) interpreted 'on its own':

(14)

$$
\begin{array}{|c|}
\hline
x, y \\
\hline
\text{Lewis}(x) \\
\text{born.in.Belfast}(x) \\
\textit{Chronicles.Narnia}(y) \\
\boxed{\begin{array}{c} z \\ \hline \text{he}(z) \end{array}} \quad \text{wrote}(z,y) \\
\hline
\end{array}
$$
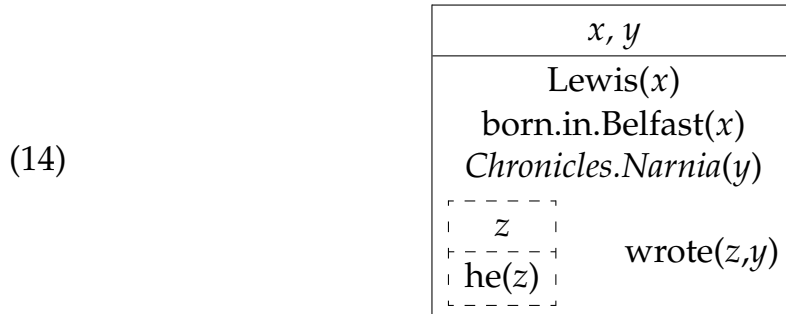
As (14) shows, a new discourse referent is added for the new NP introduced in the discourse (i.e., '*The Chronicles of Narnia*'). Moreover, the pronoun 'he' in (13) triggers the presupposition that there is a masculine entity (denoted by the dashed box) (cf. van der Sandt (1992)) and we update the DRS with the information that this masculine entity wrote y. Anaphora resolution in DRT involves equating discourse referents introduced by anaphora with appropriate and accessible[12] discourse referents in a way that leads to a maximally coherent final output DRS. For instance, we resolve the presupposition in (14) by replacing all occurrences of z with x:

(15)

$$
\begin{array}{|c|}
\hline
x, y \\
\hline
\text{Lewis}(x) \\
\text{born.in.Belfast}(x) \\
\textit{Chronicles.Narnia}(y) \\
\text{wrote}(x,y) \\
\hline
\end{array}
$$

---
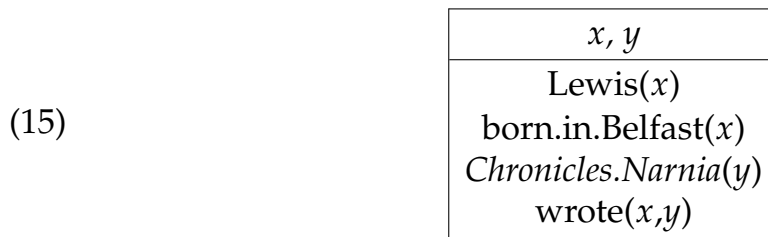
[12]Intuitively, a discourse referent is accessible if it is equally or less deeply embedded in the DRS (see Geurts et al. (2016) for specifics). For instance, x would not be accessible to y if (13) followed a statement such as (i) because x would be embedded under a negation:

(i) No writer was born in Belfast.

(ii)

$$
\begin{array}{|c|}
\hline
\\
\neg \boxed{\begin{array}{c} x \\ \hline \text{writer}(x) \\ \text{born.in.Belfast}(x) \end{array}} \quad \boxed{\begin{array}{c} z \\ \hline \text{he}(z) \end{array}} \quad \text{wrote}(z,y) \\
\\
\hline
\end{array}
$$
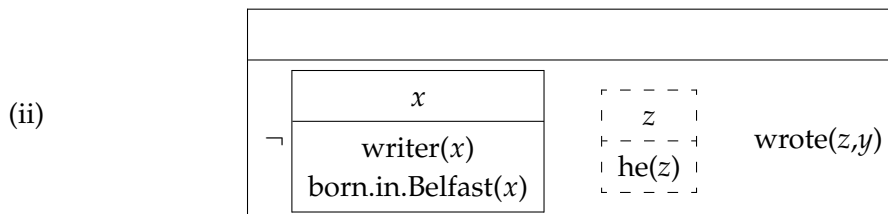
The information expressed by the entire discourse is thus that there is an entity x that is called Lewis, there is an entity y that is called *The Chronicles of Narnia*, x was born in Belfast and x wrote y.

DRS's are thus representations of the information expressed in a particular discourse as it unfolds. Hence DRT includes a 'middle level' between language and the world; discourse determines certain DRS structures and these in turn can be true about or 'verified by' the world. DRS's are verified in a model M by an 'embedding function' f from discourse referents to individuals. For instance, DRS (15) is verified by f in M iff the domain of f includes at least x and y and according to M, $f(x)$ is called Lewis and was born in Belfast, $f(y)$ is called *The Chronicles of Narnia* and $f(x)$ wrote $f(y)$.

Unofficial common grounds in DRT

Complete common grounds, containing an official common ground and (several) unofficial common grounds, can also be represented with DRS's. For instance, the complete common ground between Tolkien and myself updated with Tolkien's assertion (1) and fictional statement (2) from *The Lord of the Rings* (*lotr*) is represented as follows:

$$(16) \quad \left\langle \quad \boxed{\begin{array}{c} x \\ \hline \text{Lewis}(x) \\ \text{born.in.Belfast}(x) \\ \dots \end{array}}^{\text{Official}} \quad , \quad \boxed{\begin{array}{c} x \\ \hline \text{hobbit}(x) \\ \text{live.in.hole}(x) \\ \dots \end{array}}^{\text{Unofficial}^{lotr}} \quad \right\rangle$$
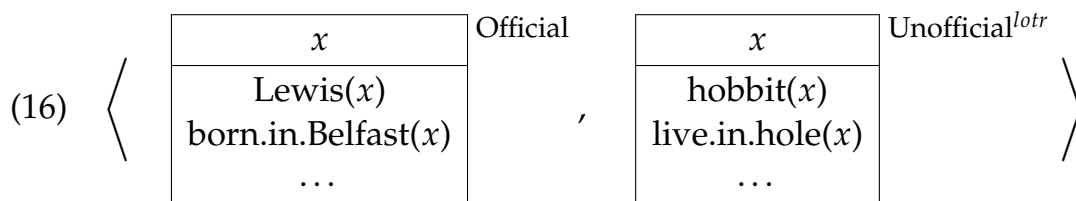
Here I assume that the official common ground between Tolkien and his reader already contained all sorts of background information before the discourse started (e.g., that water is H_2O and that Paris is the capital of France) represented by the '...'. Because a new unofficial common ground is a copy of the current official common ground, this information is also included in the unofficial common ground for *The Lord of the Rings*.

3.4 Two conflicting intuitions

Now that I have presented the basic motivations for and formalisations of the unofficial common ground accounts, I will turn to criticizing the accounts in this section. I will argue that the accounts fail to account for two

prima facie conflicting intuitions concerning the temporary acceptance of fictional truths.

3.4.1 Permanent or temporary unofficial common grounds

To grasp the following discussion, it is important to keep in mind a crucial difference between Stokke's (2013; 2018) and Eckardt's (2014) frameworks. In Stokke's account, unofficial common grounds are essentially temporary and are contrasted with "more permanent, 'official', common grounds" (Stokke, 2013, p.53). Unofficial common grounds exist only for the purpose and duration of the fictional discourse. Because we engage with many different fictions over the course of our lives, a typical complete common ground will thus consist of one permanent official common ground and several temporary unofficial common grounds that last as long as the relevant fictional discourse lasts.

On the other hand, Eckardt does not discuss whether unofficial common grounds are temporary or permanent. On the simplest construal of Eckardt's theory, which I will assume in what follows, unofficial common grounds are just as permanent as the 'normal' official common ground, i.e., once created, unofficial common grounds continue to exist alongside the official common ground and interlocutors can continue to switch between official and unofficial common grounds in subsequent discourse. Usually, a complete common ground will thus consist of one official common ground and an ever-growing number of coexisting unofficial common grounds that continue to be accessible.

3.4.2 Fictional and parafictional discourse

Analysing unofficial common grounds as essentially temporary allows Stokke to account for the intuition that the acceptance of fictional truths is temporary. Intuitively, I for example only *momentarily* accept that hobbits exist for the purpose of reading *The Lord of the Rings*. Once I stop engaging with the fictional discourse, I no longer accept this. However, there is another, prima facie conflicting, intuition that we want to account for when modelling our engagement with fiction: There is a strong sense in which I *do* retain some information about fictional content *after* engaging with the narrative. This is what allows me to continue properly interpreting fictional discourse even after a break. For instance, even after taking a break

in reading *The Lord of the Rings*, I am able to correctly interpret (17) when I pick up the novel again:

(17)   Gollum [...] held aloft the ring.

I can correctly interpret (17) because I know what Ring Tolkien is referring to and who Gollum is. Somehow, this information is still common ground between (and hence accessible to) the people that engaged with the novel. Moreover, it allows me to engage in a discussion of what is true in a particular fiction I have engaged in. For instance, even after reading *The Lord of the Rings*, I do remember that Bilbo is Frodo's cousin and would correct someone who stated otherwise. For instance, I could end up in the following discussion:

(18)   *Anne*: Did you know that Frodo from *The Lord of the Rings* was adopted by his uncle?
Merel: What? That's not true. Bilbo is Frodo's cousin.

So, after reading *The Lord of the Rings*, although I no longer accept or imagine the content that I entertained while engaging with the fictional statements of the narrative, I do not forget it. Hence I can engage in, in Recanati's (2018) terminology, 'parafictional discourse' that is based on what I have read.[13] It is important to clearly distinguish parafictional statements from fictional statements, i.e., statements that are part of a fictional narrative. Both fictional and parafictional statements can provide us with information about what is true in some fiction. But whereas fictional statements *determine* what is true in the fiction (e.g., the fact that (2) is part of *The Hobbit* makes it true in *The Hobbit* that a hobbit lived in a hole in the ground), parafictional statements *report* on what is true in the fiction; they are statements about what is true in or according to some fiction that are *not* a part of the original fictional story, but rather feature in communication about the content of a particular fictional work.

Parafictional discourse can be 'implicit' or 'explicit' depending on whether the statements include a fiction operator such as 'In story *s*,' or 'According

---

[13]Theorists employ several different (sometimes conflicting) terminologies for these and related statements. For instance, fictional statements have also been called 'fictive statements' (Currie (1990)), 'textual statements' (Zucchi (2001)) and 'authorial diktats' (Ninan (2017)). Parafictional statements have also been dubbed 'metafictive statements' (Currie (1990)) (not to be confused with 'metafict*ional* statements', see chapter 6), 'paratextual statements' (Zucchi (2001)) and 'contensive statements' (Ross (2012)).

to story *s*,'.[14] For instance, my answer in (18) is an implicit parafictional statement:

(19)   Bilbo is Frodo's cousin.

Alternatively, I could have responded to the question in (18) with explicit parafictional statement (20) or (21):

(20)   In *The Lord of the Rings*, Bilbo is Frodo's cousin.

(21)   According to *The Lord of the Rings*, Bilbo is Frodo's cousin.

Intuitively, parafictional statements – unlike fictional statements – are not just fictionally true but *really* true or false depending on actual states of affairs, i.e., the content of the relevant fictional stories. The fact that Tolkien's novel *The Lord of the Rings* was written in a certain way makes my statements (19), (20) and (21)[15] actually true and makes a statement like (22) actually false:

(22)   In *The Lord of the Rings*, Frodo is Bilbo's grandmother.

Importantly, a sentence like (2) that is found in *The Lord of the Rings*, could also function as an implicit parafictional statement if it were used in a discussion on the content of *The Lord of the Rings*. This shows that whether a utterance is a fictional statement or an implicit parafictional statement is largely a matter of context; the same sentence can function as a fictional statement (when found in a fictional work) or as an implicit parafictional statement (when found in a discussion on the content of the fictional work). Arguably, there is even a sense in which one can construe sentences that are not a *verbatim* part of a fictional narrative, but that *are* part of what is relevantly entailed by the story and entertained by the reader *while* engaging with a narrative, as an (implicit) continuation of the fictional discourse (see Semeijn and Zalta (2021)). For instance, I may think (and maybe even mutter)

---

[14]As Sainsbury (2014) notes, parafictional discourse can also feature other fiction operators such as 'partial fiction operators' like 'In/According to the first three chapters of *s*' or fiction operators such as 'It is argued in/clear by *s* that'. Following Voltolini (2019), I take these to be derivative of the 'In story *s*' or 'According to story *s*'-operators.

[15]In fact, in chapter 7 I will argue – contra consensus – that the 'In story *s*' and 'According to story *s*' operators require separate semantic analyses. The proposed analyses imply that whereas a parafictional statement such as (20) is really true, a parafictional statement such as (21) is strictly speaking false. Until chapter 7 I will largely ignore this discussion and focus mostly on parafictional statements with 'In story *s*'-operators since I take these to be the primary fiction operators.

something of the form (19) *while* reading *The Lord of the Rings*, even though (19) is never explicitly stated in the novel. Arguably, such a 'statement' can be construed as part of the fictional discourse since it takes place within the initial pretence of the novel.

Semantic analysis

One of the central objectives of a semantics of fiction is to provide a semantic analysis of parafictional statements that takes into account the different functions of fictional versus parafictional discourse and that can explain why parafictional statements such as (19), (20) and (21) ring true, whereas a statement such as (22) rings false. There is no consensus on what the appropriate semantic analysis of parafictional discourse is. Much of the current debate centers around the question of whether parafictional discourse is 'internal' or 'external' to the fiction. Discourse is internal when we describe the fictional world from a perspective within the fiction (e.g., talking about Frodo's quest to destroy the Ring as if it really took place). Discourse is external if we describe the fictional world from a perspective outside of the fiction (e.g., talking about Frodo's quest to destroy the Ring as part of the fictional events described in a fictional story). In other words, the debate is on whether parafictional statements constitute an unofficial extension of *fictional discourse* (and the pretence involved in this discourse), or whether parafictional statements are essentially a kind of non-fictional discourse, i.e., modalized assertions about the content of a fictional narrative.

According to theorists such as Everett (2013), implicit parafictional statements such as (19) constitute an (unofficial) extension of the original pretence initiated by Tolkien. Anne and myself talk about Frodo and Bilbo as if they really existed. However, if I had responded to Anne's question in (18) with the explicit parafictional statement (20), I would have adopted an external perspective to talk about *The Lord of the Rings* and hence made a type of assertion. Thus implicit parafictional discourse receives an essentially different semantic analysis from explicit parafictional discourse.

Most theorists, however, treat implicit and explicit parafictional statements on a par. There is no relevant semantic difference between answering Anne's question in (18) with either an implicit or an explicit parafictional statement: (19) is simply an abbreviation of (20).[16] These authors can be

---

[16]Admittedly, the terminology used in this dissertation is biased towards this analysis.

subdivided into two main camps: First, authors such as Recanati (2018) and Evans (1982) consider parafictional discourse in general to constitute an extension of the original fictional discourse. Both (19) and (20) are truth-valueless continuations of the *The Lord of the Rings*-pretence and hence constitute internal discourse.[17] By contrast, authors such as Currie (1990), Zucchi (2017) and Ninan (2017) claim that, whereas engaging in fictional discourse requires an internal perspective and involves pretence, neither implicit nor explicit parafictional statements involve pretence. Parafictional statements in general are simply modalized or hedged assertions about actual states of affairs in the world (i.e., statements about the content of a particular work of fiction).[18]

3.4.3 Parafictional updates on unofficial common grounds

I take both an Eckardt-style and Stokke's unofficial common ground accounts to treat parafictional discourse on a par with fictional discourse, i.e., as operating on unofficial common grounds. For instance, both (19) and (20) are analysed as proposals to update the *The Lord of the Rings* unofficial common ground. Eckardt does not explicitly discuss parafictional discourse but does describe unofficial common grounds as representations of the content of the fictional stories. Moreover, an Eckardt-style type of unofficial common ground account – where unofficial common grounds are non-temporary – is ideally suited to model the possibility of parafictional discourse: The *The Lord of the Rings* unofficial common ground does not evaporate but remains available for updating even after reading *The Lord of the Rings* precisely because we retain information about what is true in some fiction (and can discuss this) even after our engagement with it. Stokke explicitly discusses a mini-discourse that is most naturally analysed as implicit parafictional discourse.[19] He discusses the statement

---

[17]Although, for Recanati, parafictional discourse does contain an irreducible external component.

[18]As Zucchi (2017) points out, this is not to say that a sentence such as (19) could never be used in a game of pretence (e.g., I could write a fan-fiction story that is parasitic on Tolkien's narrative and that features (19)). It's just that when (19) features in a discussion on the content of *The Lord of the Rings* – when there is a clear sense that it is a true statement somehow and when it can be replaced by the explicit parafictional statement (20) – then it is an assertion.

[19]Interestingly, Stokke suggests in a footnote (Stokke, 2018, p.74) that his account of fictional statements updating unofficial common grounds is compatible with Lewis' (1978)

(23) Hobbits have hairy feet. (Stokke, 2013, p.55)

in answer to the question "Who has hairy feet?" and analyses it as a proposal to update the unofficial common ground related to *The Lord of the Rings*. This statement can be interpreted as an implicit parafictional statement (i.e., as part of a discussion on *The Lord of the Rings*) because there is a clear sense in which it is a true statement that could felicitously be replaced by the explicit parafictional statement:

(24) In *The Lord of the Rings*, hobbits have hairy feet.

Moreover, on either an implicit or explicit variant of the discourse it would be perfectly reasonable to reply: "Okay, true... But let's not talk about *The Lord of the Rings* right now."

In other words, the unofficial common ground accounts seem to fit the Recanati/Evans analysis of parafictional discourse: Both fictional and parafictional statements update unofficial common grounds and are hence treated on a par.[20] This is problematic for a number of reasons. First, there are independent reasons to prefer the Currie/Zucchi/Ninan analysis, related to

---

analysis of fictional statements as being covertly prefixed by fiction operators of the form 'In *s*,'. Here we thus have further evidence that Stokke sees no relevant difference between fictional and parafictional discourse. However, as will become clear in section 4.2.2, Lewis' semantic analysis only applies to parafictional statements. Extending it to fictional statements would have the unintuitive consequence of predicting that all sentences in *The Lord of the Rings* are true.

[20]Because Stokke only discusses implicit parafictional statements, an analysis where explicit parafictional statements operate on the official common ground as assertions (cf. Everett (2013)) is also possible in his account. In fact, yet another construal of Stokke's account is possible. In the present discussion I take Stokke's example sentence (23) to be an example of a true implicit parafictional statement. However, the example is not completely univocal; we can imagine the same discourse being placed in a context where (23) is part of a pretend conversation parasitic on *The Lord of the Rings*. Then (23) is a fictional statement, we *cannot* unproblematically replace it by (24), and we would expect a response such as "That's right! I saw one the other day". If we adopt this reading of (23), the option to analyse parafictional statements as modalized updated of the official common ground is still open to Stokke. (Strictly speaking this option is also open to Eckardt who does not explicitly discuss parafictional discourse). However, on such a version of Stokke's unofficial common ground account we still run into issues with temporality described below: whether (23) is analysed as an implicit parafictional or a (parasitic) fictional statement, both types of discourse can occur decades after engaging with a fiction and hence Stokke still requires unofficial common grounds to be continuously accessible.

the behaviour of indexicals (see Zucchi (2017)). More importantly, the Recanati/Evans view does not allow us to ascribe truth values to parafictional statements: They are, like fictional statements, not really true or false but rather fictionally true or false. However, as for instance Currie (1990) and Zucchi (2017) argue, we intuitively *do* want to maintain that (19) and (20) are true statements.

Aside from these independent theoretical reasons to not adopt the Recanati/Evans analysis, this particular analysis leads to additional issues in the unofficial common ground accounts. Both an Eckardt-style view and Stokke's account run into difficulties in accounting for the prima facie conflicting intuitions of the temporary acceptance of fictional truth on the one hand, and the fact that we do somehow retain fictional content after engaging with fictional narratives (and can hence engage in parafictional discourse) on the other hand. On the Eckardt-style view, unofficial common grounds are non-temporary (i.e., they continue to be accessible after engaging with the fictional narrative) and hence (although we can account for the possibility of parafictional discourse) we cannot account for the first intuition that fictional truths are only accepted temporarily. In Stokke's framework, unofficial common grounds are essentially temporary. However, to account for the occurrence of parafictional discourse such as (23), Stokke qualifies this by saying that "an unofficial common ground need not be temporary in the sense of lasting a short time. There are arguably common grounds that we all make use of from time to time, which are unofficial in the sense that we all know that the information they contain is fictional, or the like, but which nevertheless continue to be operative for a very long time." (Stokke, 2013, p.55). In other words, in order to account for parafictional discourse (as operating on unofficial common grounds), Stokke has to admit that unofficial common grounds remain operative long after engaging with the fictional narrative. But in what sense are such unofficial common grounds still temporary if they remain accessible after engaging with a fictional narrative (as in an Eckardt-style theory)? Hence Stokke runs into difficulties trying to account for both intuitions described above, ending up with unofficial common grounds that are both essentially temporary and continuously operative.

3.5 Conclusions and next steps

This chapter has introduced Eckardt's and Stokke's unofficial common ground accounts that distinguish an official common ground that is updated by assertions from unofficial common grounds that are updated by fictional statements. I have argued that both versions of the account run into difficulties concerning the temporariness of unofficial common grounds. Part of the difficulty is that both accounts seem to assume an analysis of parafictional statements as akin to fictional statements (i.e., as proposals to update or create an unofficial common ground). This leads to an incentive to construe unofficial common grounds as temporary and an incentive to construe them as non-temporary.

An obvious strategy to improve the unofficial common ground accounts would be to adopt the Currie/Zucchi/Ninan analysis of parafictional discourse and analyse parafictional statements as proposals to update the official common ground with hedged propositions of the form 'In/According to story s, ϕ'. If we combine this with a story of how interlocutors can derive previous unofficial common grounds based on hedged parafictional information in the official common ground (and thus continue with interrupted fictional discourse), we would be able to construe unofficial common grounds as truly temporary, i.e., only existing for the purpose and duration of the fictional discourse. As will become clear in the next chapter, the workspace account resembles this proposed improved version of the unofficial common ground accounts.

4 The workspace account

This chapter is a rewritten and significantly expanded version of sections 4 and 5 of 'A Stalnakerian analysis of metafictive statements' in *Proceedings of 21st Amsterdam Colloquium*. Parts of section 4.2.1 are adapted from 'Interacting with fictions: The role of pretend play in Theory of Mind acquisition' in *Review of Philosophy and Psychology*. The most substantial differences between this chapter and the proceedings paper include: First, an expansion of the discussion of Matravers' theory of fiction interpretation (section 4.2.1) and Lewis' analysis of parafictional discourse (section 4.2.2). Second, the inclusion of a formalisation of the workspace account in DRT (section 4.4.2). Third, the expansion of the discussion of fictive opening (section 4.5). Fourth, the addition of two suggestions for possible extensions of the account (section 4.6).

4.1 Introduction

In this chapter I introduce a novel Stalnakerian approach to address the basic puzzle of the study of the semantics of fiction, i.e., to model how content expressed by fictional discourse is quarantined from non-fictional content. The proposed account – the 'workspace account' – takes inspiration from the previously discussed unofficial common ground accounts (see chapter 3), Matravers' (2014) theory of fiction interpretation and Lewis' (1978) analysis of parafictional statements.

 This chapter starts with a brief introduction of the theoretical ingredients of the account that haven't previously been discussed: Matravers' theory (section 4.2.1) and the Lewisian fiction operator (section 4.2.2). I will then discuss the basic ideas of the workspace account (section 4.3) and offer a formalisation of them (section 4.4). I will argue that the proposed account avoids the difficulties associated with the unofficial common ground accounts (see previous chapter). It can account for the intuition that fictional

truths (e.g., that there is hobbit named Frodo) are only accepted temporarily *and* for the intuition that we do retain information about fictional truth somehow after engaging in fictional discourse, which allows us to engage in parafictional discourse (section 4.4.3) and properly continue with the fictional discourse after a break (section 4.5). I end this chapter with a brief discussion of two possible extensions of the workspace account: A version of the account where fiction that mentions Napoleon is de re about Napoleon (section 4.6.1) and an analysis of export of fictional truths as analogical reasoning (section 4.6.2).

4.2 Theoretical ingredients

Before discussing the main ideas of the workspace account, it will be useful to introduce two of its key inspirations: Matravers' theory of fiction interpretation (section 4.2.1) and Lewis' analysis of the fiction operator (section 4.2.2).

4.2.1 Matravers' two stage model

The workspace account is inspired by Matravers' (2014) theory of fiction interpretation. Matravers follows Friend (2008; 2011a; 2012) in criticizing the widely adopted 'consensus view' (Walton (1990); Currie (1990)) that draws a sharp distinction between fiction interpretation and non-fiction interpretation, i.e., whereas nonfictional truths are to be believed, fictional truths are to be 'imagined'. He argues that the characterizations of the cognitive attitude of imagination that are on offer (e.g., Currie (1990); Meskin and Weinberg (2006)) apply equally to non-fiction as well. For instance, simulationists characterize imagination as 'running mental states offline', which indicates an absence of direct perceptual inputs and of a motivation to act. So, when reading *The Hobbit* I use my imagination because I have no direct perceptual inputs of Bilbo and no incentives to act upon the described events. However, Matravers points out that when I read a non-fictional article about Donald Trump in *The New York Times*, I also have no direct perceptual inputs of Trump nor a direct motivation to act and hence also 'run mental states offline'. Therefore no special link between fiction and imagination is established.

According to Matravers, the fundamental flaw in the consensus view is its confusion of the distinction between engaging with fiction and engaging with non-fiction, with the more fundamental and cognitively primary distinction between engaging with 'confrontation situations' and engaging with 'representation situations'. In confrontation situations, people have a direct possibility to act because their mental states are caused by perceptual inputs from objects in their immediate surroundings (e.g., a situation in which a tiger enters your house and you have the possibility to run and shout for help). In representation situations, by contrast, people have no direct possibility to act because their mental states are caused by mere representations of objects (e.g., a situation in which someone tells you about a tiger that entered your house yesterday). Engaging with either a fictional or a non-fictional narrative is simply an example of being in a representation situation (where you have no direct possibility to act). Hence, at least our primary engagement with fictional narratives (e.g., reading *The Lord of the Rings*) involves essentially the same cognitive processes as engaging with nonfictional narratives (e.g., reading a biography such as Monk's *The Duty of Genius*). Whether a narrative is fictional or non-fictional, when we read or listen to it, we simply entertain its content by building a representation or 'mental model' based on the incoming discourse.[1] Hence, insofar as there is a role for the concept of 'imagination' in Matravers' framework, it is something that is at play for both fiction and non-fiction.

Matravers does not discard the distinction between fiction and non-fiction entirely. In his 'two stage' model of narrative interpretation only the first stage (i.e., entertaining a narrative's content) is neutral with regards to fictionality. In the second stage the distinction between fiction and non-fiction becomes apparent: "Put very roughly, simulating fiction scenarios does not result in our forming beliefs and simulating non-fictional scenarios does result in our forming beliefs." (Matravers, 2014, p.27). In other words, while reading *The Hobbit* may involve the same cognitive processes as reading

---

[1]In fact, there is a debate on whether 'merely entertaining' content is even possible (see Recanati (forthcoming-a, forthcoming-b) for a recent discussion). On 'cancellation accounts', understanding a proposition implies accepting it as true. This acceptance can subsequently be cancelled (e.g., in case we engaged with fiction). On a cancellation version of the workspace account, fictional and non-fictional discourse uniformly update the stable belief-based common ground directly. Fictive closure would be a cancellation operation through which the updated stable common ground is 'cancelled' and added to the previous stable common ground under the relevant fiction operator.

The Duty of Genius, the reader will have a disposition to believe the entertained content in the case of the biography, and a disposition not to believe the entertained content in the case of the fictional novel. Importantly, for Matravers this is a *rough* characterization of the fiction/non-fiction divide; we cannot base a definition of fictional or non-fictional narratives on the presence or absence of a disposition to believe its content. What prevents us from doing so is the fact that fictional narratives can also contain content that we are disposed to believe (e.g., I may be inclined to learn facts about 19th century etiquette in England from reading *Pride and Prejudice*). However, at present my main aim is to define fictional and non-fictional *statements*. Such definitions do not imply anything about whether fictional *narratives* can also contain non-fictional statements or vice versa. I construe Matravers' theory as allowing for a sharp distinction here: we have a disposition to believe the content of non-fictional statements (or assertions) and a disposition not to believe the content of fictional statements.[2]

Before moving on it will be useful to highlight an issue the reader may see in Matravers' theory at this point. After engaging with a fictional narrative (e.g., after entertaining thoughts about Frodo and the Ring), you simply have a disposition *not* to believe the entertained content. So fictional content, unlike non-fictional content, just 'evaporates' after reading or listening to some story. This is not satisfactory. As was discussed in the previous chapter, there are two prima facie conflicting intuitions that a theory of fictional and non-fictional discourse should explain. Matravers' theory is able to account for the intuition that fictional truths are only accepted temporarily but it does not account for the intuition that people somehow retain knowledge about what is true in some fiction after engaging with it. After reading *The Lord of the Rings* I probably don't *believe* that Frodo was born in the Shire, but I do retain this information in some (quarantined) way. Part of the appeal of the consensus view is that it can account for this latter intuition. The cognitive attitude of imagination is supposed to function in a way that is parallel to the cognitive attitude of belief. So, after reading *The Lord of the Rings*, I do not *believe* that Frodo was born in the Shire, but this is still a part of my *imagination* (or simulated belief) based on *The Lord of the Rings*.

As discussed in the previous chapter, unofficial common ground accounts that render unofficial common grounds permanent do account for the sec-

---

[2]In section 4.6.2 I suggest a possible strategy to develop a Matraversian account of fictional and non-fictional narratives.

ond intuition. Ideally, we would have a theory that can preserve this benefit of such a theory without inheriting the difficulties discussed. The workspace account is an attempt at formulating a Stalnakerian theory that takes the best of both worlds and thus can account for both described intuitions. As will become clear below, the main 'fix' is to add to Matravers' framework a disposition to believe *parafictional* content based on the entertained content after engaging with fictional discourse. The next section introduces Lewis' analysis of parafictional discourse which I adopt as part of the workspace acccount.

4.2.2 The Lewisian analysis of the 'In s'-operator

I adopt Lewis' (1978) analysis 2 of the 'In story s'-operator. Lewis treats this operator as an intensional operator, i.e., as quantifying over possible worlds. Below is a simplified representation of this operator's semantic definition:

> "In s, ϕ" is true iff in all possible worlds compatible with s, ϕ is true

The obvious question now is what makes a world "compatible" with some fiction. In his seminal paper 'Truth in Fiction' (1978), Lewis goes through several analyses of this notion.

First, we cannot simply take worlds compatible with s to be worlds where the plot of s is enacted. A basic problem with this analysis is that the actual world could be one of those worlds but in the actual world s is a *fictional* narrative. This relates to Kripke's (1980) well received point that, even if – purely by coincidence – our world turned out to have included someone named 'Sherlock Holmes' who solved crimes in 19th century London, still the name 'Holmes' as used by Arthur Conan Doyle would *not* refer to this person. Doyle wrote his novels as pure fiction and never met this real-life Holmes. Hence it is false in our world (one of the worlds where the plot of the Sherlock Holmes novels is enacted) that the name 'Holmes' in s refers to someone. But surely it should come out true in the world of the fiction that the name 'Holmes' in s refers to someone!

Lewis argues that to overcome these difficulties we must consider a fictional narrative not as an abstract set of propositions but as something that involves an act of story-telling (and hence a story-teller). We thus arrive at Lewis' first attempt at an analysis of truth in fiction:

> **Analysis 0**: "In s, ϕ" is true iff in all possible worlds where s is told as known fact (rather than fiction), ϕ is true

Although all worlds where *s* is told as known fact are worlds where the plot of *s* is enacted, the actual world is not amongst these worlds. Hence we avoid the difficulties described above.

Analysis 0 gets basic facts concerning *explicit* truth in fiction right. For instance, it is explicitly stated in the Sherlock Holmes novels that Holmes smokes a pipe and so this is true in all worlds where the story is told as known fact. Hence the analysis rightly predicts that it is true in the Sherlock Holmes novels that Holmes smokes a pipe. However, Analysis 0 disregards *implicit* fictional truths, i.e., things that are not explicitly stated in the fiction but that we nevertheless consider to be true in it. For instance, it is true in the Sherlock Holmes novels that water is H_2O, that whales are mammals and that "Holmes does not have a third nostril" (Lewis, 1978, p.41). Such fictional truths are 'imported' into the fiction as part of the background information that we assume when engaging with a fiction. However, none of these things is actually stated in the Sherlock Holmes novels, nor do they follow from what is explicitly stated. This means that there are possible worlds where the Sherlock Holmes stories *are* told as known fact but where the above things are false! Hence it is not true in *all* worlds compatible with the Sherlock Holmes novels that water is H_2O, that whales are mammals and that Holmes does not have a third nostril. Implicit fictional truths therefore do not come out as true in the fiction on Analysis 0.

Lewis argues that in order to incorporate background information into the analysis fictional truth, we have to analyse the fiction operator as a counterfactual, i.e., what is true in *s* is what would be true if *s* were told as known fact. In other words, we take the actual world as our 'starting point' and see what it would be like if *s* were told as known fact in our world. On Lewis' (1973) analysis of counterfactuals, a statement of the form 'If ϕ, then ψ' is true iff some possible world where ϕ and ψ are true is closer to the actual world than any world where ϕ is true but ψ is not true. A world is 'closer' to some other world if it is more similar to it. Hence, a counterfactual 'If ϕ, then ψ' is true iff ψ is true in all ϕ-worlds that are closest (most similar) to the actual world. Application of this analysis of counterfactuals to the case of fiction gives us Analysis 1:

> **Analysis 1**: "In *s*, ϕ" is true iff in all possible worlds where *s* is told as known fact that are closest to the actual world, ϕ is true

Worlds where the Sherlock Holmes novels are told as known fact *and* water is H_2O are closer to the actual world than worlds where the Sherlock Holmes

novels are told as known fact and water is *not* H_2O. Hence, Analysis 1 correctly predicts that implicit fictional truths are true in the fiction. Lewis thus incorporates a version of the Reality Principle (see section 3.3.1) into his analysis of fictional truth, i.e., when engaging with fiction assume the fictional world to be as much like our own world as the text allows.

Last but not least, Lewis discusses one final complication; Analysis 1 makes little-known and even unknown facts relevant to fictional truth. On Analysis 1, whatever is actually the case will also be true in the fiction (unless explicitly contradicted by it). But, given that it is the case that Trump won the elections in 2016, is it then also true in the Sherlock Holmes novels that Trump wins the elections in 2016? Admitting this would force us to admit that what is true in a fiction is subject to constant change (as what is *really* true is subject to constant change). Whether or not we want to allow (the constant) importation of such 'remote' and irrelevant fictional truths is a matter of debate (see e.g., Friend (2017)). However, as Lewis argues, sometimes little-known facts can be detrimental to the plot (as it was envisioned by the author) of a fictional narrative. This can lead to counterintuitive results. For instance, in *The Adventure of the Speckled Band*, Holmes claims to have solved a murder case by showing that someone has been killed by a viper that climbed up a bell rope. Gans (1970) has argued that, since vipers cannot actually climb ropes, either it's true in *The Adventure of the Speckled Band* that the snake reached its victim some other way, or Holmes has not solved the case at all. This is not intuitive; Holmes is always right!

If we want to resist Gans' conclusions (i.e., want our analysis to predict that it is true in *The Adventure of the Speckled Band* that Holmes was right about the viper climbing the bell rope), truth in fiction should not depend on little-known facts. Doyle (and his readers) didn't realize that vipers cannot climb ropes and hence this doesn't come out as true in the fiction. On Lewis' final analysis 2, fictional truth depends on what was general common belief when the fiction was written:

> **Analysis 2**: "In s, ϕ" is true iff in all possible worlds where s is told as known fact that are closest to the community of origin's overt conception of the actual world, ϕ is true

Here the "community of origin's overt conception of the actual world" consists in the overt beliefs about the actual world in the community of origin of the relevant fiction. The 'overt beliefs' of a community are the

beliefs that are generally and openly shared, i.e., general common beliefs. Analysis 2 allows us to still import truths such as that water is H_2O. This is overt belief in the community of origin of the Sherlock Holmes novels, i.e., generally common belief between Doyle and his readers. However, information such as that Trump wins the election in 2016 or that vipers cannot climb ropes will not be imported into the fiction. Although a single reader of the Sherlock Holmes novels may personally believe either of these things (and it may even become overt belief after the fiction has been published), it is *not* part of the community of origin's overt beliefs about the actual world and hence *not* part of what is true in the Sherlock Holmes novels.

4.3 Workspaces

Now that I have introduced its main theoretical ingredients, I will turn to introducing the workspace account. I will start with an informal discussion of the account's key components in this section and then turn to formalisations in section 4.4.

The workspace account incorporates a concept similar to Stokke's (2013; 2018) temporary unofficial common ground: A 'workspace'. As in Stokke's unofficial common ground account, a distinction is drawn between the stable (official) common ground and a temporary common ground. The stable common ground between any group of people contains their shared presuppositions concerning actual states of affairs. This common ground can (for now) be construed as belief-based (e.g., it is stable common ground between Tolkien and myself that Paris is the capital of France because we both believe this, believe that the other believes this, etc.). It is 'stable' in so far as content that enters the common ground remains common ground persistently unless and until possible belief revision forces us to revoke it (e.g., Tolkien and I persistently believe that Paris is the capital of France).

The workspace is a temporary common ground that contains all shared presuppositions between a speaker and hearer while engaging with some specific discourse. In line with Matravers' theory of narrative interpretation, the starting assumption of the workspace account is that our primary engagement with a fictional discourse (e.g., reading or listening to a story) involves the same interpretative processes as engaging with a non-fictional discourse. Hence, unlike Stokke's unofficial common grounds, the workspace

is neutral with respect to fictionality.[3] I assume that what is part of the workspace between a speaker his hearers at the onset of a new fictional or non-fictional discourse is a copy of the current stable common ground. For non-fiction this just embodies the central tenets of Stalnakerian context dependence and presupposition satisfaction. For instance, I can start a new non-fictional discourse and felicitously assert something about terrorist attacks in Paris because it is stable common ground between us what Paris is. Hence the workspace will also contain this information and hence my assertion will be interpretable to my interlocutor. For fiction this amounts to an implementation of the Reality Principle, i.e., that when we engage with a fiction we understand it against a background or importation of factual information about the actual world. This workspace is then updated with the propositions that are expressed in the discourse, i.e., the propositions that are entertained (or used in 'mental model building') by speaker and hearers while engaging in the discourse.[4] For instance, while reading *The Lord of the Rings* it is temporarily common ground between Tolkien and myself that Frodo was born in the Shire. Likewise, while reading Monk's biography of Wittgenstein *The Duty of Genius*, it is temporarily common ground between Monk and myself that Wittgenstein was Austrian. The workspace is construed as acceptance-based; when updating the workspace we are merely entertaining this content, not believing it. Thus Tolkien and I temporarily commonly *accept* that Frodo was born in the Shire and Monk and I temporarily commonly *accept* that Wittgenstein was Austrian, i.e., we both temporarily accept this while engaging with the discourse, temporarily believe that the other accepts this, etc.[5]

---

[3]The term 'workspace' is also used by Nichols and Stich (2000) in their cognitive approach to pretend play. Their account incorporates a 'Possible World Box' which is a 'workspace' (that is kept separate from our beliefs) in which our cognitive system builds and temporarily stores representations of possible worlds. However, unlike my account's workspace, this possible world box is only operational in the case of pretend play (e.g., fictional discourse) and related tasks. Non-fictional interaction (e.g., non-fictional discourse) operates on the belief box.

[4]A similar idea is developed in Kamp's (2018) mentalistic framework. Kamp introduces a compartment (K_{dis}) for the neutral place where we build representations of the content of the current discourse before forming judgements about the truth of the propositions expressed by K_{dis}.

[5]Actually, in these non face-to-face cases, hearer and speaker do not simultaneously engage in the discourse and hence do not have the relevant (temporary) attitudes simultaneously. See section 2.3.1.

Unlike the stable common ground that exists permanently and unlike Stokke's unofficial common grounds, the workspace is truly temporary. It remains in use and existence solely for the purpose and solely for the duration of a specific conversation. In other words, speaker and hearer only accept the propositions expressed by a particular discourse as long as they are engaging in it. For instance, as soon as I stop reading *The Lord of the Rings* and hence stop entertaining the propositions expressed by it, I stop accepting the propositions of the discourse. Hence the content of the workspace evaporates and it stops to exist. Thus we can speak of one single workspace coming into existence and disappearing again (or, alternatively, becoming active/accessible and non-active/inaccessible again) rather than speaking of several independent unofficial common grounds for different fictions.

Assertions and fictional statements are defined as proposals to update the workspace and stable common ground in a three-step algorithm where, conform Matravers' theory, the first two steps are uniform for fiction and non-fiction. The first step is opening a temporary workspace alongside the stable common ground at the start of the discourse. The second step is updating this workspace with the content of the (fictional or non-fictional) discourse. The workspace that came into existence with the first update (i.e., with the first proposition we are entertaining) remains accessible and in existence during subsequent updates caused by the same, possibly multi-sentence, discourse. In other words, when entertaining propositions from some narrative (e.g., *The Lord of the Rings* or *The Duty of Genius*), a workspace is created with the first update and we continue to further update this workspace with subsequent assertions or fictional statements. When speaker and hearer stop entertaining propositions from this discourse (i.e., as I stop reading or listening), the workspace loses its content and evaporates. As I subsequently engage in a new discourse (e.g., I start chatting to my neighbour or start reading the *Harry Potter* series), I again update, and thereby activate, a new workspace (with possibly different interlocutors).

In the third and final step of the algorithm the difference between assertions and fictional statements becomes apparent. The difference consists in how, as soon as the discourse ends, the quarantined content in the workspace is brought back to update on the stable common ground. I propose two distinct closure operations on workspaces; what differentiates assertions from fictional statements is whether the relevant speech act is a proposal to,

at the end of the possibly multi-sentence discourse, perform 'assertive' or 'fictive closure'.

If we were to simply transfer Matravers' theory to a Stalnakerian framework this would entail that in the case of assertive closure the content of the workspace is believed – and hence added to the stable common ground – and in the case of fictive closure it is not. However, to do justice to the intuition that fictional content does not simply evaporate after engaging in fictional discourse, the workspace account includes a parafictional update at fictive closure. The entertained propositions are retained as being true '*in* the fiction'. More specifically, after engaging with a story s, the content of the workspace ϕ is added to the stable common ground under the relevant Lewisian fiction operator, i.e., as parafictional information of the form 'In story s, ϕ'.[6] So, even though after reading *The Lord of the Rings*, you do not believe that Frodo was born in the Shire (and this does not become common ground between you and Tolkien), you *do* believe the embedded statement that in *The Lord of the Rings*, Frodo was born in the Shire and this becomes common ground between you and Tolkien.[7]

Placing the content of the workspace under Lewis's Analysis 2 fiction operator at fictive closure fits neatly with the assumption that a new un-updated workspace is a copy of the current stable common ground.[8] The propositions that are true in worlds where some fiction s is told as known fact that are most similar to the community of origin's overt conception of the actual world coincide with the propositions that are part of the stable common ground between an author and his readers that is updated with propositions expressed by s.[9] For instance, it is part of the workspace while reading the Sherlock Holmes novels that water is H_2O. This information was stable common ground between Doyle and myself (and his other readers)

---

[6] Interestingly, in their cognitive approach to pretend play Nichols and Stich (2000) come to a similar conclusion. Their account involves a belief update of the form 'if pretence premise p were true, then q', where q is the content in the possible world box. They use this update to explain how people come to exhibit pretend behaviour.

[7] The proposed account of the common ground updates caused by fictional discourse resonates with some elements of Nichols' and Stich' (2000) account of the cognitive structures required for pretend play (see footnotes 6 and 3). See Semeijn (2019) for a Matraversian account of pretend play and its role in Theory of Mind development.

[8] See also Zucchi (forthcoming) who combines the notion of a Stalnakerian common ground with the Lewisian fiction operator for a similar purpose.

[9] I thus abstract away from 'hearers' that do not share the overt beliefs of the relevant community of origin.

and is hence copied into the workspace at the first fictional update. Likewise, it is part of the workspace while reading the Sherlock Holmes stories that Holmes smoked a pipe since this is part of the propositions expressed by the fiction and hence the workspace is updated with this information. Last but not least, it is *not* part of the workspace while reading the Sherlock Holmes novels that Trump won the election in 2016. This information was never stable common ground between Doyle and myself (and his other readers), nor was it part of the fictional narrative and hence it does not enter the workspace.

In sum, assertions are defined as proposals to open and update a workspace and as a result of that trigger assertive closure. Fictional statements are defined as proposals to open and update a workspace and as a result of that trigger fictive closure. In other words, whether we are engaging in fictional or non-fictional discourse, the propositions expressed are temporarily common ground (i.e., part of the workspace) while engaging in the discourse. At the end of the discourse this temporarily accepted content becomes stable common ground. In the case of non-fiction it is added to the common ground directly. In the case of fiction it is added to the common ground embedded under the relevant fiction operator. Fictional content is thus effectively quarantined; firstly in the temporary workspace and secondly embedded under a fiction operator.

4.4 Formalisation

I will now provide two formalisations of the three step algorithm involved in assertions and fictional statements. First, by representing common grounds as sets of propositions (section 4.4.1) and second by representing common grounds in DRS's (section 4.4.2).

4.4.1 Sets of propositions

Opening up a workspace

The first step when engaging in a new discourse is opening up or bringing to existence a new workspace. This step is uniform for fiction and non-fiction. Both the first assertion of a non-fictional discourse and the first fictional statement of a fictional discourse will trigger this. I model this step as an

operation on an ordered pair consisting of a stable common ground (*C*) and the empty set, resulting in an ordered pair of the same stable common ground and a new workspace that is active and accessible for updating:

$$\langle C, \varnothing \rangle + p = \langle C, C \rangle + p$$

As in the unofficial common ground accounts (see previous chapter), I assume that what is common ground between a speaker and his hearers at the onset of a new fictional discourse is a copy of the current stable (official) common ground. Hence, the new unupdated workspace (created in the first step of the algorithm) is a copy of the current stable common ground (*C*).

Updating the workspace

Once we have our initial workspace (*W*) set up, we start updating with the incoming information in the second step of the algorithm. This step is also uniform for fiction and non-fiction: fictional statements and assertions alike update the workspace with the propositions that they express:[10]

$$\langle C, W \rangle + p = \langle C, W * p \rangle$$

The workspace that came into existence with the first statement of a particular discourse is subsequently updated by the statements that follow and that are part of the same discourse.

Assertive and fictive closure

As discussed, what differentiates assertions from fictional statements is how, at the end of the discourse, they update the stable common ground. Assertions trigger assertive closure, fictional statements trigger fictive closure. In the representations below, both closure operations take an ordered pair $\langle C, W \rangle$ containing a stable common ground and an updated, active workspace, and return an ordered pair with a new stable common ground and the empty set.

In the case of assertive closure, the updated workspace is adopted as the new stable common ground. Because a new workspace is a copy of

---

[10]As in the formalisation of the unofficial common ground accounts (see section 3.3.1), we assume that the $*$ operator unionizes sets if they are consistent and otherwise resolves the inconsistency appropriately.

the current common ground, asserting a proposition p thus boils down to updating the stable common ground C to $C * p$ (as in the traditional Stalnakerian framework):

$$\text{Assertive closure: } \langle C, W \rangle \rightarrow \langle W, \varnothing \rangle$$

Fictive closure returns an ordered pair in which the updated workspace is added to the original stable common ground as parafictional information, i.e., under the relevant fiction operator. In the formalisation below I assume that we are keeping track of n fictions $(1, ..., n)$. The 'In fiction i'-operator ($\square_i$) takes as its argument the proposition $\cap W$, which is the intersection of the propositions in W (i.e., the information that corresponds to how we have interpreted the story) and it gives the set of worlds in which it is true that worlds compatible with i are those in $\cap W$. Note that the fiction operator $\square_i$ is thus taken semantically in this context, i.e., as a function from sets of possible worlds to sets of possible worlds:

$$\text{Fictive}^i \text{ closure: } \langle C, W \rangle \rightarrow \langle C \cup \{\square_i(\cap W)\}, \varnothing \rangle$$

So after engaging with some fictional discourse i, the stable common ground will contain parafictional information concerning i and the workspace evaporates. Because we normally engage in different fictional discourses (and know other people to do so as well), a typical stable common ground between any group of people will contain (apart from information about the actual world) parafictional information about several distinct fictions under different 'In fiction i'-operators. In this sense there are in fact multiple different fictive closure operators related to different fictional works.

4.4.2 Discourse representation structures

Below, I illustrate the updates on workspaces and stable common grounds triggered by assertion (1) and Tolkien's fictional statement (25) taken from *The Lord of the Rings* in the box notation of DRT:[11]

(1)   C.S. Lewis was born in Belfast.

(25)   Frodo had a very trying time that afternoon.

---

[11]See also Maier (2017) and Kamp (forthcoming) for modern semantic implementations of the consensus view, in a rather different, purely mentalistic version of DRT.

In order to do so we require a version of the Lewisian fiction operator that operates on DRS's:

> For any DRS K and any story s, $\Box_s K$ is a well-formed DRS condition and $\llbracket \Box_s K \rrbracket^{f,w} = 1$ iff in all possible worlds w' compatible with $f(s)$, $\llbracket K \rrbracket^{f,w'} = 1$

For example, a DRS may contain a DRS-condition of the form $\Box_{lotr} K$ where a sub-DRS K is embedded under the *Lord of the Rings* fiction operator. This condition may then be verified in a model or not. For instance, if a main DRS contains a sub-DRS that is embedded by the *Lord of the Rings* fiction operator and that sub-DRS contains the information that Frodo was born in the Shire, then the main DRS is verified if it is true in *The Lord of the Rings* that Frodo was born in the Shire.

Assertions

First, a simplified representation of assertion (1) defined as a proposal to open and update a workspace and, as a result of that, perform assertive closure. First, a new workspace is opened which is a copy of the current common ground (in this case it already contains, amongst other things, the information that the sun rises in the east):

$$(26) \quad \left\langle \boxed{\begin{array}{c} C \\ \hline \text{sun.rises.in.east} \\ \dots \end{array}} , \varnothing \right\rangle$$

$$(27) \quad \left\langle \boxed{\begin{array}{c} C \\ \hline \text{sun.rises.in.east} \\ \dots \end{array}} , \boxed{\begin{array}{c} W \\ \hline \text{sun.rises.in.east} \\ \dots \end{array}} \right\rangle$$
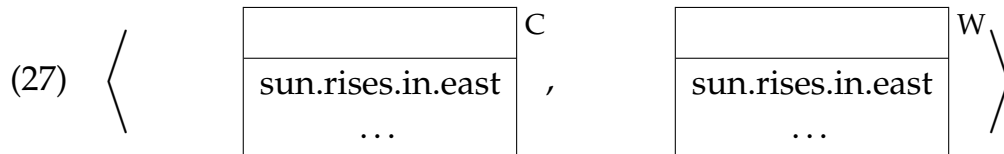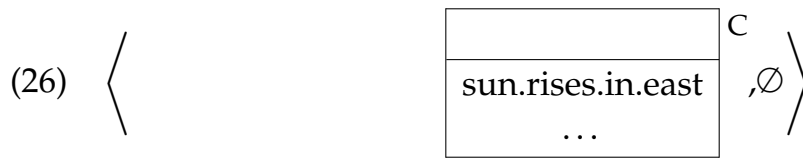
This workspace is then updated with the proposition expressed by (1). I use primed discourse referents (x', y', ...) to represent the content of workspaces:

$$(28) \quad \left\langle \boxed{\begin{array}{c} C \\ \hline \text{sun.rises.in.east} \\ \dots \end{array}} , \boxed{\begin{array}{c} W \\ \hline x' \\ \hline \text{sun.rises.in.east} \\ \dots \\ \text{Lewis}(x') \\ \text{born.in.Belfast}(x') \end{array}} \right\rangle$$
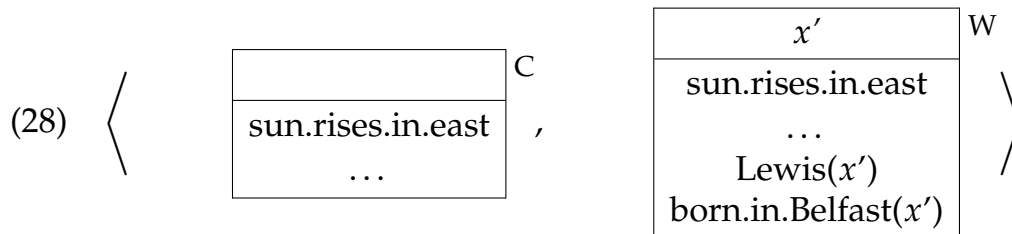
As soon as the non-fictional discourse ends, assertive closure is triggered, i.e., the content of the common ground is replaced by the content of the workspace, leaving us with a new common ground that contains the information expressed by (1) and an evaporated workspace:

$$
(29) \quad \left\langle
\begin{array}{|c|}
\hline
\multicolumn{1}{|r|}{x'} \\
\hline
\text{sun.rises.in.east} \\
\ldots \\
\text{Lewis}(x') \\
\text{born.in.Belfast}(x') \\
\hline
\end{array} \; C
\quad , \varnothing \right\rangle
$$
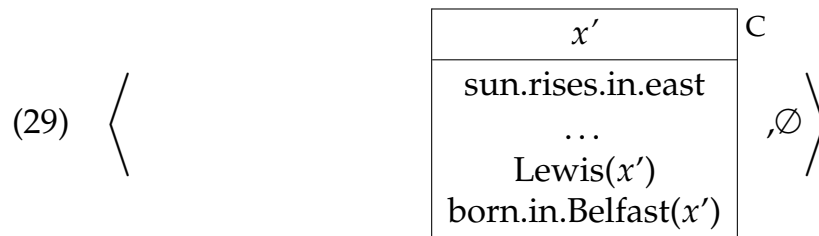
The net result of this algorithm is thus exactly the same as for standard DRT updating on the common ground without opening and closing workspaces. The payoff of adding a workspace update lies in the way it allows us to model the essential similarities and dissimilarities between non-fiction and fiction.

Fictional statements

Next, a simplified representation of fictional statement (25) defined as a proposal to open and update a workspace and as a result of that perform fictive closure. Again, a new workspace is opened up which is a copy of the current common ground:

$$
(30) \quad \left\langle
\begin{array}{|c|}
\hline
\multicolumn{1}{|r|}{} \\
\hline
\text{sun.rises.in.east} \\
\ldots \\
\hline
\end{array} \; C
\quad , \varnothing \right\rangle
$$

$$
(31) \quad \left\langle
\begin{array}{|c|}
\hline
\multicolumn{1}{|r|}{} \\
\hline
\text{sun.rises.in.east} \\
\ldots \\
\hline
\end{array} \; C
\quad ,
\begin{array}{|c|}
\hline
\multicolumn{1}{|r|}{} \\
\hline
\text{sun.rises.in.east} \\
\ldots \\
\hline
\end{array} \; W
\right\rangle
$$

This workspace is then updated with the proposition expressed by (25):

$$
(32) \quad \left\langle
\begin{array}{|c|}
\hline
\multicolumn{1}{|r|}{} \\
\hline
\text{sun.rises.in.east} \\
\ldots \\
\hline
\end{array} \; C
\quad ,
\begin{array}{|c|}
\hline
\multicolumn{1}{|r|}{x'} \\
\hline
\text{sun.rises.in.east} \\
\ldots \\
\text{Frodo}(x') \\
\text{trying.time}(x') \\
\hline
\end{array} \; W
\right\rangle
$$
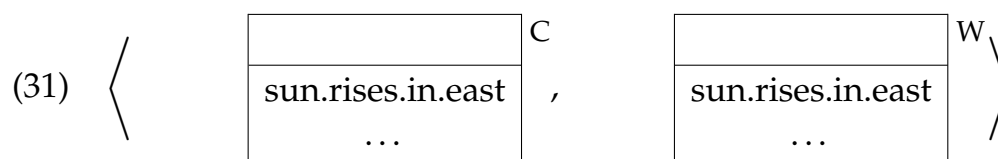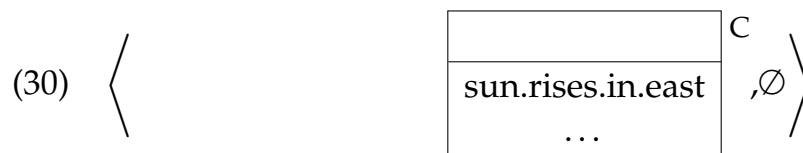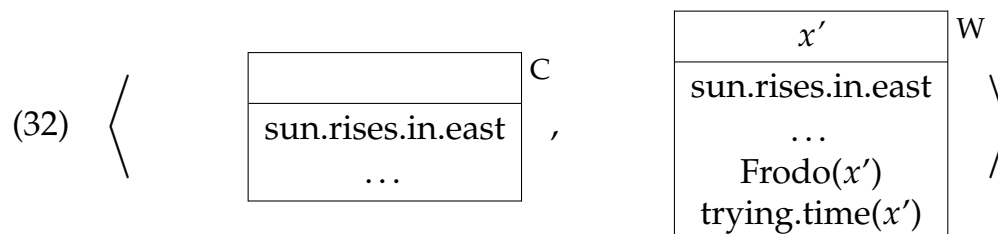
As soon as the fictional discourse ends, fictive closure is triggered, i.e., the content of the workspace is added to the common ground as parafictional information under the relevant fiction operator. This leaves us with a new common ground that contains the information that in *The Lord of the Rings* the proposition expressed by (25) is true, and an evaporated workspace:

(33)
$$
\left\langle \quad \Box_{lotr} \boxed{\begin{array}{l} \text{C} \\ \hline \text{sun.rises.in.east} \\ \quad \ldots \\ \boxed{\begin{array}{l} x' \\ \hline \text{sun.rises.in.east} \\ \quad \ldots \\ \text{Frodo}(x') \\ \text{trying.time}(x') \end{array}} \end{array}} \quad ,\varnothing \right\rangle
$$
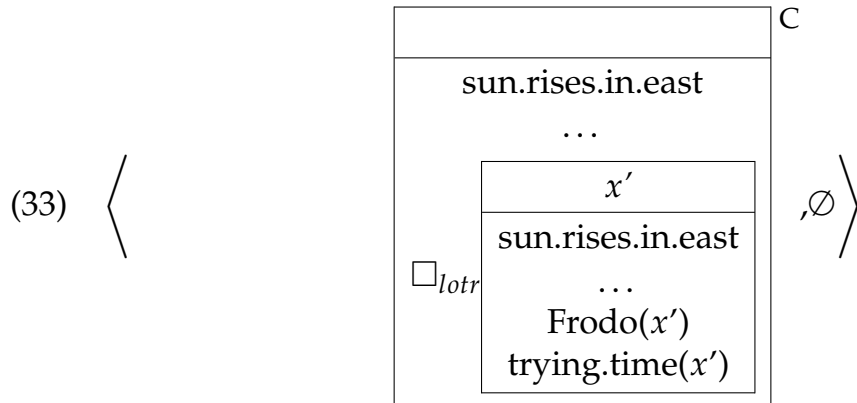
4.4.3 Parafictional discourse

As we have seen, the parafictional update at fictive closure operates on the stable common ground, i.e., a hedged or modalized proposition becomes part of the stable common ground. Likewise, if we engage in a discussion on the content of some fiction (e.g., *The Lord of the Rings*) and someone utters an implicit or explicit parafictional statement such as (19) or (20), these update the stable common ground with a hedged proposition:

(19)   Bilbo is Frodo's cousin.

(20)   In *The Lord of the Rings*, Bilbo is Frodo's cousin.

In line with the Currie/Zucchi/Ninan analysis of parafictional discourse, parafictional statements are analysed as modalized assertions about actual states of affairs (i.e., the content of particular novel) and hence are proposals to update the workspace with a modalized proposition and trigger assertive closure.

Concretely, any arbitrary parafictional proposition p consists of an 'In fiction i'-operator related to some fiction i ($\Box_i$), and some proposition (q): $p = \Box_i q$. In the earlier set theoretic representation we can thus represent the updates caused by parafictional statements by substituting p for $\Box_i q$ in the three-step algorithm for assertions. After opening up a new workspace we

thus update that workspace with $\Box_i q$:

$$\langle C, W \rangle + \Box_i q = \langle C, W * \Box_i q \rangle$$

At the end of the (possibly multi-sentence) parafictional discourse about the content of fiction i, we perform regular assertive closure: $\langle C, W \rangle \rightarrow \langle W, \emptyset \rangle$. The updated workspace is adopted as the new stable common ground (which now also contains $\Box_i q$).

As an illustration we can represent the updates caused by (20) (or (19)) in DRT. After opening up a workspace (that is a copy of the current common ground), the hedged parafictional proposition updates the workspace directly, i.e., it is temporarily common ground that in *The Lord of the Rings*, Bilbo is Frodo's cousin:

(34)

$$\left\langle \quad \boxed{\begin{array}{c} \\ \hline \cdots \end{array}}\,C \quad , \quad \Box_{lotr}\boxed{\begin{array}{c} \cdots \\ \hline \boxed{\begin{array}{c} x', y' \\ \hline Frodo(x') \\ Bilbo(y') \\ cousin(x',y') \end{array}} \end{array}}\,W \quad \right\rangle$$
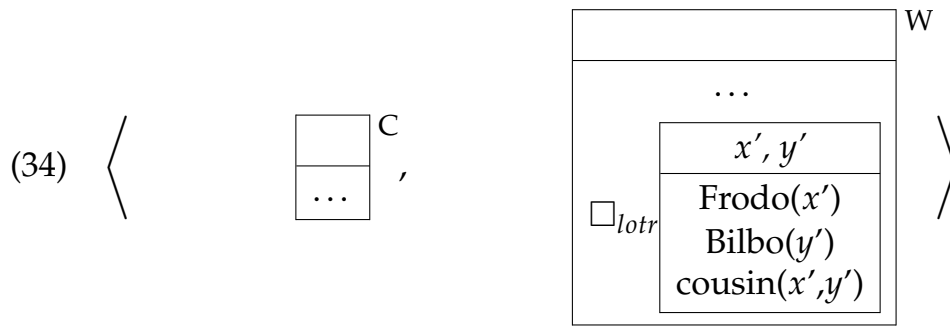
At the end of the discourse assertive closure is triggered, i.e., the common ground is replaced by the current workspace:

(35)

$$\left\langle \quad \Box_{lotr}\boxed{\begin{array}{c} \cdots \\ \hline \boxed{\begin{array}{c} x', y' \\ \hline Frodo(x') \\ Bilbo(y') \\ cousin(x,y) \end{array}} \end{array}}\,C \quad ,\emptyset \right\rangle$$
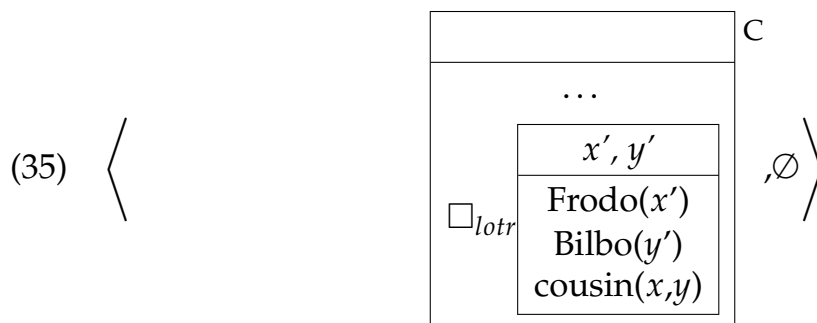
Hence the stable common ground now also contains the parafictional information that in *The Lord of the Rings*, Bilbo is Frodo's cousin.

There are thus basically two ways to update the common ground with parafictional information. Either you engage in fictional discourse s (e.g., reading *The Lord of the Rings*) or you engage in parafictional discourse about the content of s (e.g., engaging in a discussion about the content of *The Lord of the Rings*). Either discourse results in updates of the stable

common ground with parafictional information concerning *s*. The intuitive difference between the two processes lies in what kind of propositions you entertain (i.e., update your workspace with) during the discourse; whether you entertain propositions such as "Bilbo is Frodo's cousin" or propositions such as "In *The Lord of the Rings*, Bilbo is Frodo's cousin".

The analysis of parafictional discourse as hedged assertions allows me, unlike Stokke (2013; 2018) and Eckardt (2014), to ascribe truth-values to parafictional statements such as (19) and (20) just as we do with regular assertions. Moreover, the workspace account avoids the difficulties with unofficial common ground accounts that treat fictional and parafictional discourse on a par. In the workspace account, after engaging in a fictional narrative such as *The Lord of the Rings*, the fictional content of our workspace evaporates; we accept propositions such as that wizards exist only temporarily. However, this and other fictional content does not evaporate completely. After the fictional discourse it becomes part of the stable common ground as parafictional information. Thus, after engaging in *The Lord of the Rings* it is stable common belief that (in *The Lord of the Rings*,) Bilbo is Frodo's cousin. This explains how we can, after engaging in a fiction, engage in parafictional discourse. When engaging in parafictional discourse, we make regular assertions that rely on, and update the common ground with, hedged propositions and therefore no 'permanent' unofficial common ground or workspace related to *The Lord of the Rings* is called for.

However, the workspace account does not yet explain how the parafictional update of the stable common ground allows us to continue with some fictional discourse after taking a break. I turn to this issue in the next section.

4.5 Fictive opening

4.5.1 Picking up where we left off

Up until this point I have presented a *tabula rasa* interpretation of fiction where nothing is common ground between speaker and hearer about what is true in the fiction before starting to engage with it (except that the fiction – any fiction – conforms to the common ground based version of the Reality Assumption). Obviously this isn't always the case when engaging with fictional narratives. Most importantly, people can take breaks while engaging with fictional discourse. For instance, I may have read *The Lord of the Rings*

yesterday, entertained some of its content, stopped engaging with it (and hence performed fictive closure), bookmarked the page where I stopped reading, engaged in all sorts of other non-fictional and fictional discourse, and today pick up the book where I left off. In this case it is already common ground between Tolkien and myself that some things (e.g., that there is a hobbit named Frodo) are true in *The Lord of the Rings* before I start to engage with it (again).[12]

A feature of the workspace account, as it is presented above, is that after engaging in a fictional narrative and entertaining its content, all that we are left with in our common ground is parafictional information. Hence, if we would – when returning to our fictional discourse after a break – create a new workspace by making a copy of the current stable common ground, we would start 'from scratch'. All previously introduced discourse referents for fictional objects would become inaccessible because they are embedded under a fiction operator in the common ground and hence also in the new workspace. It thus becomes unclear how I could for instance interpret a fictional statement such as (17) when getting back to *The Lord of the Rings*:

(17)   Gollum [...] held aloft the ring.

What ring is Tolkien referring to? Who is Gollum? If the content of the previous updates caused by the fictional discourse of *The Lord of the Rings* is inaccessible, I cannot answer these questions. In order to account for such anaphoric links, the new workspace will have to contain all the propositions that were included in the original workspace just before fictive closure (e.g., a description of some unique ring). Hence the workspace account is in need of some further mechanism to explain how we are able to retrieve the the final state of the relevant earlier fiction workspace as our current active workspace.[13]

---

[12]Conversely, while writing *The Lord of the Rings,* Tolkien most probably took numerous breaks. Whenever he would return to the writing table and continue the fictional discourse it would already be common ground between him and his readers that certain things are true in *The Lord of the Rings*.

[13]This problem does not arise with breaks in non-fictional discourse because with assertive closure we adopt the updated workspace as the new stable common ground. Hence, when continuing in a non-fictional narrative, the new workspace will contain (at least) all propositions that were included in the original workspace. Moreover, this problem does not arise in unofficial common ground accounts that construe unofficial common grounds as non-temporary.

A possible solution is to claim that, apart from adding parafictional propositions to the stable common ground, fictive closure also involves retaining a copy of the updated workspace (which is adopted as new workspace when continuing in the same narrative). This results in a theory resembling Stokke's (2013; 2018) and Eckardt's (2014) accounts (see chapter 3), involving (something akin to) unofficial common grounds. However, this move invites the problems concerning temporality associated with the unofficial common ground accounts; if we maintain that readers of *The Lord of the Rings* hold onto a *The Lord of the Rings*-workspace containing propositions such as that wizards exist, we no longer account for the intuition that we accept such fictional propositions only temporarily.

A more promising solution is to introduce a mechanism that explains how we, when continuing in a familiar narrative, fill in our workspace based on the available parafictional propositions in the common ground. At first sight this seems like a straightforward task. In fictive closure you copy the updated workspace in its entirety and add it to the stable common ground under an 'In fiction *s*'-operator. So, when you continue to engage in a fictional narrative after taking a break, you simply reverse the fictive closure, i.e., perform 'fictive opening': identify the relevant fiction-operator and copy everything that is under this operator to the workspace.

The need for a fictive opening mechanism shows why the DRT box syntax – which brings out the occasionally criticized 'representational character' of DRT – is not only a visually efficient tool but also theoretically relevant. It is not possible to define a fictive opening operation if we represent common grounds as sets of propositions (which in turn are sets of possible worlds). The difficulty with this formalism is that the individual propositions in the common grounds are presented as having no structure. When we perform fictive closure we update the stable common ground with parafictional information and thus simply add a new set of possible worlds (e.g., we add the set of worlds in which in the novel *The Lord of the Rings*, Frodo is a hobbit, Frodo inherits the Rings, etc.). In other words, there is no 'parafictional *The Lord of the Rings*' marker in the stable common ground and hence no straightforward mechanism to select the appropriate propositions to perform fictive opening.

4 The workspace account

We therefore need a framework that does place such a structure on the information in the common ground. DRT does exactly this.[14] Information in the common ground is represented in DRS's which are complex structures involving fiction operator conditions on DRS's. These fiction operator conditions can function as parafictional markers that allow us to select the appropriate propositions to perform fictive opening. So, for instance, when opening up a workspace after taking a break in reading *The Lord of the Rings*, the content under the *The Lord of the Rings* fiction operator in the stable common ground is adopted as new workspace:

(36)

$$
\left\langle
\begin{array}{|l|}
\hline
 \hspace{3cm} \text{C} \\
\hline
\text{sun.rises.in.east} \\
\ldots \\
\Box_{lotr}
\begin{array}{|l|}
\hline
x \\
\hline
\text{sun.rises.in.east} \\
\ldots \\
\text{Frodo}(x) \\
\text{trying.time}(x) \\
\hline
\end{array} \\
\hline
\end{array}
\quad , \quad
\begin{array}{|l|}
\hline
x' \hspace{1.5cm} \text{W} \\
\hline
\text{sun.rises.in.east} \\
\ldots \\
\text{Frodo}(x') \\
\text{trying.time}(x') \\
\hline
\end{array}
\right\rangle
$$

---
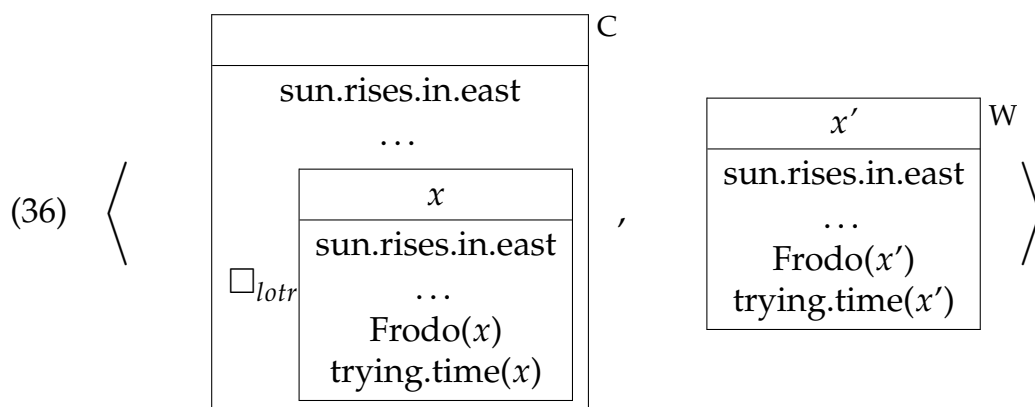
[14] An alternative framework that also places the necessary structure on propositions and hence allows for parafictional markers, is the so-called 'structured propositions' framework (see for instance Soames (1985) and Cresswell (1985)). Propositions are not sets of possible worlds, but complex entities with a structure similar to the sentences that expresses them and with constituents that carry the semantic values of expressions occurring in these sentences. For example, in Soames' neo-Russellian approach the sentence "Scott does not run" expresses the following proposition:

$$< NEG, << s >, R >>$$

Here, s is Scott, R is the property of running and NEG is the truth function for negation. Thus, the negation operator is a distinct constituent of the proposition expressed. We can analyse the 'In fiction i'-operator in parafictional statements in a similar fashion. The following is a simplified representation of the proposition expressed by parafictional statement (20):

$$< \Box_{lotr}, << b, f >, C >>$$

Here, b is Bilbo, f is Frodo, C is the property of being someone's cousin and $\Box_{lotr}$ is the 'In *The Lord of the Rings*'-operator. In this way we place structure on propositions that allows for parafictional markers and hence enables us to perform fictive opening.

Hence we can re-create the last known state of the workspace, K, from a parafictional condition of the form $\Box_s K$, generated in the stable common ground after fictive closure.

4.5.2 Genre conventions

Another way the common ground can contain parafictional information before engaging with a fictional narrative is through what Lewis calls 'inter-fictional carry-over' of fictional truth. Additional fictional truths may derive from prior knowledge about what is true in other fictional stories. This can for instance happen because the narrative is part of a larger canon that deals with the same fictional world (e.g., the Harry Potter book series, the Star Wars Expanded Universe), though, arguably, such narratives can also be construed as simply a continuation of the same fictional discourse. However, more general genre conventions can add fictional truths as well:[15] In a typical fairy tale about a knight going on a quest to slay a dragon, we may anticipate that the dragon breathes fire, even if that has not (yet) been stated explicitly. The question is how the information that dragons breathe fire enters the workspace since it is neither stated explicitly in the text, nor part of the stable common ground (assuming that it is common ground that there are no dragons).

I propose that genre conventions are imported in a way similar to fictional truths derived from previous engagement with a fiction, i.e., through fictive opening. As mentioned above, when I continue reading a fictional narrative after a break that triggered fictive closure, I re-create the last known state of the workspace, K, from a parafictional condition of the form $\Box_s K$ in the stable common ground. Genre expectations may be stored in the common ground in terms of parafictional conditions as well.[16] Consider the unknown fairy tale from before. I pick up the book, and on the basis of the cover picture and first few lines ("Once upon a time in a faraway land there lived a knight...") I decide that I'm dealing with a fairy tale. At the same time, it's

---

[15]Genre conventions may also influence whether fictional statements are judged reliable and how they update the workspace (see chapter 8).

[16]Here I only consider conventions related to different genres of fiction. Non-fiction 'genre conventions' (e.g., In a news report, a family taking tea at their dining-room table means that the family is 'normal') are analysed as unprefixed stereotypic knowledge in the common ground. See Matravers (2014) and Zucchi (forthcoming) for a uniform treatment of fiction and non-fiction genre conventions.

common ground that 'in fairytales, dragons breathe fire' or, in other words, that if some y is a fairytale, then it is true in y that dragons breathe fire:

$$
(37) \quad \left\langle \quad
\boxed{\begin{array}{l}
\underline{\hspace{3cm} x \hspace{3cm}} \;\; \text{C} \\
\cdots \\
\text{book}(x) \;\; \text{fairytale}(x) \;\; \text{read}(i,x) \\
\boxed{\dfrac{y}{\text{fairytale}(y)}} \Rightarrow \boxed{\Box_y \;\; \boxed{\text{dragons.breathe.fire}}}
\end{array}}
\quad , \varnothing \right\rangle
$$
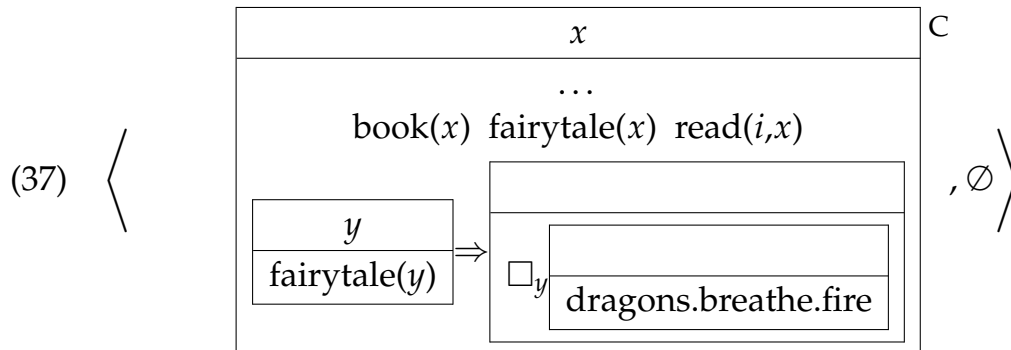
The workspace that we open to represent the story at hand should be a copy of the common ground, as usual, but also contain the information that holds in stories of this type, as stored in quantified parafictional statements like in (37). We can define the fictive opening mechanism to take care of continued reading and genre assumptions uniformly: when starting to interpret a story s, make a copy of the stable common ground and merge that with all K such that $\Box_s K$ is part of (or can be inferred on the basis of, as in (37)) the stable common ground.

4.6 Possible extensions

Now that I have presented the basics of the workspace account I will end this chapter with a brief discussion of two potential extensions of the account: a de re version of the account (section 4.6.1) and an analysis of export as analogical reasoning (section 4.6.2).

4.6.1 Fiction about non-fictional objects

What the DRT formalisation makes apparent is that, as it is formalised above, the workspace account is 'descriptivist', i.e., all content in the workspace (discourse referents and conditions) is placed under a fiction operator at fictive closure. The account thus adheres to a 'fictional substitute' analysis of fiction, i.e., fictions that make reference to non-fictional objects (e.g., historical fiction) are analysed as not being de re about a real life person or thing but rather about a fictional substitute of it. Consider Tolstoy's fiction

novel *War and Peace* that features Napoleon as one of its characters. *War and Peace* contains fictional statement (38):

(38)   Napoleon [...] sat on his small gray Arab horse a little in front of his marshals.

Assuming it is already stable common ground between Tolstoy and his audience before engaging in the fictional discourse who Napoleon was, there is a discourse referent for the real world Napoleon in the stable common ground. Alongside the rest of the common ground, this discourse referent is copied into the new unupdated workspace at the beginning of the fictional discourse:

(39)
$$
\left\langle
\begin{array}{|c|}\hline
x \\\hline
\text{Napoleon}(x) \\
\text{emperor}(x) \\
\ldots \\\hline
\end{array}{}^{C}
\; , \;
\begin{array}{|c|}\hline
x' \\\hline
\text{Napoleon}(x') \\
\text{emperor}(x') \\
\ldots \\\hline
\end{array}{}^{W}
\right\rangle
$$
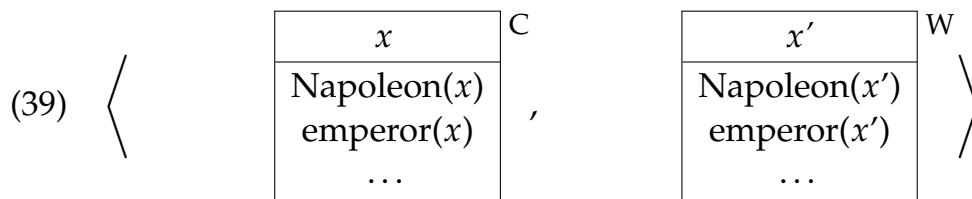
Hence a new discourse referent is added to the workspace that is separate from the discourse referent for Napoleon in the stable common ground. The workspace is then updated with the proposition expressed by fictional statement (38). At fictive closure this workspace is added to the stable common ground under the *War and Peace* fiction operator:

(40)
$$
\left\langle
\begin{array}{|c|}\hline
x \\\hline
\text{Napoleon}(x) \\
\text{emperor}(x) \\
\ldots \\
\Box_{w\&p}
\begin{array}{|c|}\hline
x' \\\hline
\text{Napoleon}(x') \\
\text{emperor}(x') \\
\ldots \\
\text{sat.on.horse}(x') \\\hline
\end{array} \\\hline
\end{array}{}^{C}
\; , \varnothing
\right\rangle
$$
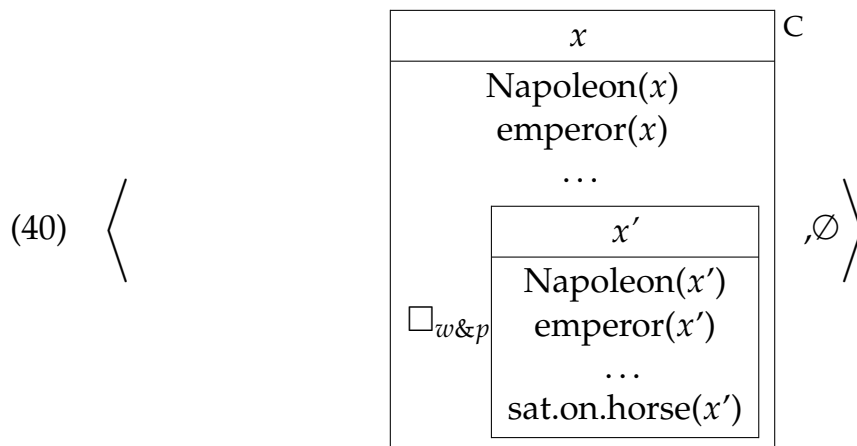
In this stable common ground there is a discourse referent x in the main DRS for the real world Napoleon and the conditions in the main DRS express that x was an emperor etc. There is also a discourse referent x' for the fictional substitute Napoleon in the embedded DRS. The conditions in the embedded DRS express that the fictional substitute of Napoleon was an emperor but

also that he sat on his small gray Arab horse a little in front of his marshals. So, after reading *War and Peace* it is stable common ground that in *War and Peace*, there is an entity that is called Napoleon, was an emperor, etc. Thus, strictly speaking, *War and Peace* is not really de re about the real world Napoleon but about a fictional substitute (i.e., something with a lot of the same properties as Napoleon – but also some fictional ones – that exists in a fictional world). The idea that fictions that make reference to non-fictional objects are strictly speaking about their fictional substitutes is in line with for instance Wieland's (forthcoming) proposal but against the dominant view that such fictions are literally de re about the real world objects they make reference to (e.g., Friend (2011b) and Maier (2017)).[17]

The workspace account can, however, be made compatible with the intuition that historical fictions are literally de re about the relevant real world objects. We can add an additional mechanism that links or anchors (cf. Maier (2017)) discourse referents that are copied to the workspace to their real counterparts in the stable common ground:

(41)
$$\left\langle \begin{array}{|c|}\hline x^1 \\\hline \text{Napoleon}(x^1) \\ \text{emperor}(x^1) \\ \ldots \\\hline\end{array}{}^{\text{C}} \, , \begin{array}{|c|}\hline x^{1\prime} \\\hline \text{Napoleon}(x^{1\prime}) \\ \text{emperor}(x^{1\prime}) \\ \ldots \\\hline\end{array}{}^{\text{W}} \right\rangle$$

At fictive closure all discourse referents in the workspace that have substitute discourse referents in the main DRS are then replaced by these non-fictional substitutes when the content of the workspace is added to the common ground embedded under the relevant fiction operator. Hence the common ground after fictive closure triggered by statement (38) will look as follows:

---

[17]As far as we consider adherence to a non-dominant view a problem, it is one that is shared by the unofficial common ground accounts. Unofficial common grounds are separated from the official common grounds and are independently updated. In order to appropriately engage with a fiction that makes reference to a non-fictional object, the relevant unofficial common ground needs to copy discourse referents and conditions on them from the official common ground. For instance, to properly interpret a fictional statement such as (38) we copy a separate discourse referent for the fictional substitute of Napoleon into the *War and Peace* unofficial common ground. See also Semeijn and Zalta (2021) who argue that explicit parafictional statements are ambiguous between a de re and a de dicto reading.

$$(42) \quad \left\langle \quad \boxed{\begin{array}{|c|} \hline x^1 \qquad \text{C} \\ \hline \text{Napoleon}(x^1) \\ \text{emperor}(x^1) \\ \dots \\ \Box_{w\&p} \boxed{\begin{array}{c} \\ \hline \text{Napoleon}(x^1) \\ \text{emperor}(x^1) \\ \dots \\ \text{sat.on.horse}(x^1) \\ \end{array}} \\ \hline \end{array}} \quad , \emptyset \quad \right\rangle$$
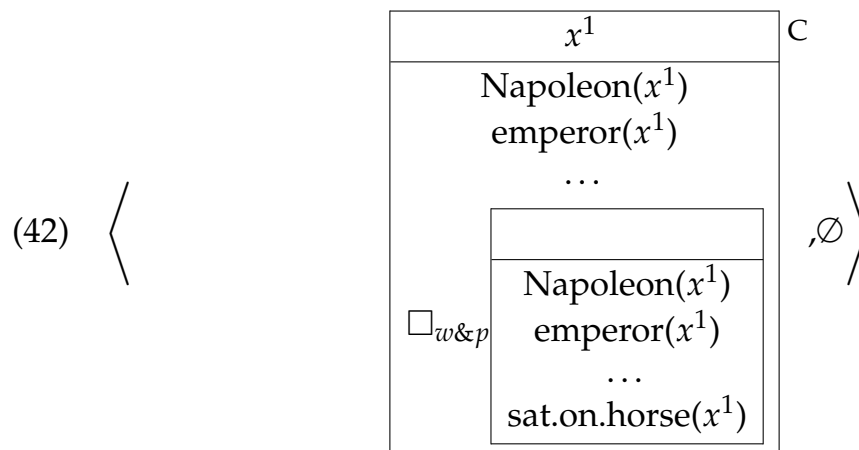
Here there is only one discourse referent for the real world Napoleon. It is common ground that he is called Napoleon and that he is an emperor. Moreover, it is common ground that in *War and Peace*, he (the real world Napoleon) was called Napoleon, was an emperor and sat on his small gray Arab horse a little in front of his marshals. Such an approach seems intuitive but does raise familiar questions concerning 'quantifying in' (see e.g., Quine (1956); Kaplan (1968)), e.g., can *War and Peace* 'know' or represent Napoleon under multiple different guises (through different acquaintance relations) and hence express contradictory de re information about him?

In the following chapters I will assume the basic descriptivist version of the workspace account. This choice will not play a role in subsequent discussions with the notable exceptions of chapter 6 and 8. In chapter 6 we will return to the issue of fiction about non-fictional objects and see that one of the central strategies to deal with 'metafictional' discourse seems to force a move away from a simple descriptivist account anyway. In chapter 8 we assume there is at least some kind of linking or anchoring between discourse referents for real life entities and their fictional counterparts.

4.6.2 Export of fictional truth as analogical reasoning

As has been discussed, a Matraversian characterization of fictional and non-fictional statements cannot straightforwardly be extended to definitions of narratives. According to such a definition, we have a disposition to believe the content of non-fictional narratives and we have a disposition not to believe the content of fictional narratives. However, fictional narratives sometimes express content towards which we have a disposition to believe. Take the following quote from Fleming's *Thunderball*:

(43) New Providence, the island containing Nassau, the capital of the Bahamas, is a drab sandy slab of land fringed with some of the most beautiful beaches in the world. (Example taken from Friend (2008))

Although it is part of a fictional narrative, we can learn empirical facts about the real world through statements such as (43) in a process that Gendler (2000) calls 'narrative as clearinghouse'. After reading *Thunderball*, I really do believe that Nassau is on the island New Providence. Hence we 'export' this truth from the fiction. To account for such dispositions we may opt for a so-called 'patchwork theory of fiction' (e.g., Currie (1990)): a fictional narrative can consist of both fictional and non-fictional statements.[18]

An alternative strategy is to go for a 'knitwork theory of narratives': whereas fictional narratives consist of fictional statements, non-fictional narratives consist of non-fictional statements. We can opt for such an account (on a Matraversian or on the consensus view) if we adopt the (by now) familiar idea that after engaging with a fiction, we update with parafictional beliefs about its content. According to such a definition, although we may not have a disposition to believe the content of a fictional narrative, we do have a disposition to believe parafictional information based on its content. In the following I will suggest that from a subset of these parafictional beliefs we can derive unprefixed beliefs through analogical inferencing. Hence people can have an indirect disposition to believe some of the propositions expressed by fictional statements.

The two-dimensional approach

This account is inspired by the fact that people talk about import and export principles of fictional truth as being based on assumed or perceived similarities between fictional worlds and the actual world; if a fictional story s is supposed to be realistic with respect to a certain cluster of facts and p is in this cluster, we can derive p from $\Box_s p$ and vice versa (cf. Ichino and Currie (2017)). We can capture this intuitive idea in terms of analogical reasoning.

To clarify the concept of analogical reasoning I adopt the terminology and schematization of Hesse's (1966) and Bartha's (2010) two-dimensional approach in which an analogical inference is justified if [1] a 'source' and 'tar-

---

[18]See Friend (2011a) and Stock (2011) who identify several issues with such patchwork theories. See also García-Carpintero (2013) who avoids patchwork problems by defining fictional and non-fictional narratives in normative terms.

get system' exhibit a 'positive analogy' (i.e., similarities) and [2] the source system exhibits a 'vertical relation' of correlation[19] between properties in and outside the positive analogy that can be extended to the target system, given that [3] there is no 'negative analogy' (i.e., dissimilarities) that is also relevantly correlated to the positive analogy and hence prohibits this. We can for instance represent (part of) Reid's (1785) analogical argument about the heavens as follows:

Source	**Target**	
P	P*	[positive analogy]
A	¬A*	[negative analogy]
Q		
	Q*	

P: Earth orbits the sun	P*: Mars orbits the sun
A: Earth's radius is 6,371 km	A*: Mars' radius is 6,371 km
Q: Earth can sustain life	Q*: Mars can sustain life

So, Reid was justified in inferring that Mars could sustain life (Q*) because Earth and Mars exhibit the positive analogy of both orbiting the sun (P ∧ P*), Earth's orbiting the sun was taken to be correlated to Earth sustaining life (P is correlated to Q) and the negative analogy exhibited by Earth and Mars (A ∧ ¬A* i.e., Mars being smaller than Earth) was not taken to be relevantly correlated to the positive analogy.

Analogical reasoning with parafictional beliefs

We can also provide a justification for reasoning about what is true about the real world (or at least about what the author thinks is true about the real world), based on what is true in a fictional world if we put export in terms of analogical inference. Suppose I have read *Thunderball* and have obtained the parafictional beliefs that 'In *Thunderball*, there exists a rebreather (a small device that allows you to breathe underwater)', 'In *Thunderball*, the Bahamas start East of the coast of Florida' and 'In *Thunderball*, Nassau is on New Providence' because these things are stated in the narrative. Suppose I already knew that '(Actually) the Bahamas start East of the coast of Florida'.

---

[19]Theorists differ in the requirements they put on these vertical relations (e.g., causal, predictive, mere correlation etc.). For the purpose of this dissertation I assume that vertical relations may consist in mere correlation.

We can represent the situation as follows:

Source	Target
Source	**Target**
P	P*
A	¬A*
Q	
---	---
	Q*

P: In *Thunderball*, the Bahamas start East of the coast of Florida

P*: The Bahamas start East of the coast of Florida

A: In *Thunderball*, there exists a rebreather

A*: There exists a rebreather

Q: In *Thunderball*, Nassau is on New Providence

Q*: Nassau is on New Providence

Because the *Thunderball* worlds and the real world exhibit the positive analogy P and P* (i.e., the fiction is realistic with respect to this geographical fact), properties P and Q are related (i.e., belong to the same cluster of geographical facts) and there is no relevant negative analogy (e.g., A is not in the same cluster of facts as P and Q), we are licensed to extend the vertical relation to the real world and hence infer B* (i.e., that '(Actually) Nassau is on New Providence'). In a similar way, we are licensed to infer that New Providence, the island containing Nassau, the capital of the Bahamas, is a drab sandy slab of land fringed with some of the most beautiful beaches in the world from reading (43) in *Thunderball* because we perceive that *Thunderball* is realistic with respect to geographical facts (i.e., Bond does not visit made up places). We can infer this despite the fact that we have perceived *Thunderball* to not be realistic with respect to facts concerning technological possibilities and advances in the 1960s, i.e., we know the author took the liberty of make up fancy but unrealistic gadgets such as the rebreather.

Horizontal relations of similarity between the fictional and the real world concerning certain clusters of facts can be strengthened by perceived similarities (as in the example above) or by genre conventions (see Ryan (1991)). For instance, even before reading a Jane Austen novel (and hence before perceiving any similarities in clusters of facts), I already expect the novel to *not* be realistic with respect to geographical facts (I know Austen's novels sometimes contain made up villages or estates), but I *do* expect the novel to be realistic with respect to facts concerning etiquette and social practices in 19th century British upper class.

An important caveat here is that these inferences (as are analogical inferences in general) are of course not deductive but always subject to uncertainty. I can only infer that probably, Nassau is actually on New Providence from reading this in *Thunderball*. Moreover, it is difficult to systematically determine what facts belong to a particular cluster of facts (cf. Norton (2021)). For instance, Austen's novels are not necessarily realistic with respect to small and unknown geographical facts, but *are* realistic with respect to facts concerning well-known, larger cities and countries (e.g., in Austen's novels, London is in the south of England and is its capital). I leave further exploration of an analogical reasoning account of export to future research. An advantage of the envisioned account, as opposed to patchwork theories, is that it is easily extendable to other media. We don't have to posit that the composer, painter or filmmaker has somehow 'asserted' certain things in order to account for the fact that we can learn facts about the real world from their creations. Moreover, the account is easily extendable to what Gendler (2000) calls 'narrative as factory' (i.e., learning general implicit truths such as 'long exposure to excessive power can corrupt even the humblest person' from engaging with a fiction such as *The Lord of the Rings*), and easily extendable to export of presuppositional content of fictional statements (e.g., from reading "Bond rode to the airport" in *Thunderball* I can infer that there was an airport on New Providence in the 1960s). Unlike a patchwork theorist, the knitwork theorist does not have to posit that these general truths or the presuppositions in fictional statements were somehow 'asserted'. We only have to establish that they were made true in the fiction and were viable for export.

4.7 Conclusions

This chapter has introduced the key components of the workspace account. Rather than using different update rules for fictional statements and assertions (as in the unofficial common ground accounts), I propose a uniform workspace update along with distinct assertive and fictive closure operations. Assertions are defined as proposals to (open and) update a temporary acceptance-based workspace and as a result of that trigger assertive closure, i.e., the workspace is adopted as the new stable common ground. Fictional statements are defined as proposals to (open and) update a workspace and as a result of that trigger fictive closure, i.e., the workspace is added to the

stable common ground under the relevant fiction operator. A new unup-dated workspace is a copy of the current stable common ground that is merged with all information embedded under the relevant fiction operators in the stable common ground. The account thus effectively quarantines fictional content (i.e., first in a temporary workspace and then embedded under a fiction operator). I have argued that the account can account for the intuition that fictional truths are only accepted temporarily (i.e., as information in the workspace) and for the intuition that we do retain information about fictional truths even after fictional discourse (i.e., as parafictional beliefs stored in the stable common ground).

The workspace account is the basic framework that is used in this dissertation and will be applied to several puzzles present in the semantics and philosophy of fiction literature. Chapters 5, 6 and 8 present possible applications and/or embellishments of the basic account.

5 Lies, bald-faced lies and parafictional updates

This chapter is a rewritten version of the unpublished paper 'Bald-faced lies and parafictional updates' which has been presented at several international conferences.

5.1 Introduction

Now that I have presented the basic framework for modelling assertions and fictional statements in terms of common ground updates, let's see how related speech acts fit into this framework. In this chapter I consider lies and so-called 'bald-faced lies'. Suppose Brian invites Alice to his birthday party but Alice does not want to attend and hence says (44) even though actually she is feeling perfectly fine. Compare this to Tolkien's fictional statement (25):

(44)   I have a cold.

(25)   Frodo had a very trying time that afternoon.

Both Tolkien and Alice defied the Gricean maxim of quality; they said something they believed to be false. Yet, there is a strong intuitive difference between these two speech acts. Whereas (44) is a lie, (25) is 'merely' a fictional statement. A natural way of phrasing the difference is to say that lying, contrary to fiction telling, involves an intention to deceive (e.g., Augustine (395); Williams (2002)); whereas Alice had the intention to deceive Brian into believing that she had fallen ill, Tolkien never meant to deceive anyone into believing that some hobbit had a trying time on some afternoon.

I would like to thank Teresa Marques, the audience at the *Speech acts and fiction* seminar at the University of Genoa and the audience at the DGfS workshop *Post-truth: the semantics and pragmatics of saying "what you believe to be false"* at the University of Bremen for valuable feedback and discussion.

A well-known counterexample to such definitions of lying is what Sorensen (2007) has dubbed the bald-faced lie, i.e., a statement that involves no intention to deceive but that we do intuitively call a lie. For instance, Keiser (2016) discusses an example found in *The Godfather II* where mafioso Pentangeli is about to testify against Corleone (the Godfather) during a senate hearing. Because of Pentangeli's previous unofficial statements, it is common knowledge in the courtroom that Pentangeli knows Corleone and his crimes. However, Corleone has Pentangeli's brother attend the hearing in order to remind Pentangeli of the fact that by testifying, he puts the safety of his family in jeopardy. When asked whether he served the Godfather, Pentangeli testifies:

(45)   I never knew no Godfather.

Pentangeli thereby ensures that Corleone is not convicted and his own family remains safe. Although (45) cannot involve an intention to deceive anyone (everyone in the courtroom knows that Pentangeli *did* know the Godfather, everyone knows that everyone knows this, etc.)[1] we do call it a (bald-faced) lie.

There are three main strategies that theorists have adopted to deal with bald-faced lies. First, we can bite the bullet: Bald-faced lies are valid examples of lies so apparently lying does not necessarily involve an intention to deceive. To make such an analysis tenable we need to provide a definition of lying that does not involve intention to deceive but somehow still distinguishes lies from fictional statements and mistakes (e.g., Carson (2006); Fallis (2009); Stokke (2013, 2018)). A second strategy is to deny the validity of the counterexample by arguing that bald-faced lies like in the Godfather scenario actually *do* involve an intention to be deceptive (Lackey (2013)). Third, we can deny the validity of the counterexample by arguing that bald-faced lies are not really lies (e.g., Meibauer (2014)), for instance because they are only quasi-assertoric (e.g., Dynel (2011); Leland (2015); Keiser (2016); Maitra (2018)).[2] In this chapter I align myself with the latter group. I incorporate an

---

[1] Arguably, this is not the case for all speech acts that we may call 'bald-faced lies'. In general, I assume that a speaker who states some proposition p that is blatantly false to only a part of their audience (e.g., a politician fabricating facts in some public statement) performs two separate speech acts: A lie towards people with whom it is not common knowledge that p is false and a 'real' bald-faced lie (cf. Maitra (2018); Keiser (2016)) towards people with whom this *is* common knowledge.

[2] Cf. Harris (forthcoming) who argues that because bald-faced lies are not assertions, not all lies are assertions.

analysis of lies and bald-faced lies into the workspace account according to which bald-faced lies function like fictional statements. The obvious benefit of this approach is that we can avoid the "unhappy divorce" (Lackey (2013); see also Keiser (2016)) of definitions of lying and intention to deceive. One of the central contributions of this chapter is to argue that the bald-faced liar's aim of 'going on the record' – sometimes taken to show that bald-faced lies are assertions (by e.g., Sorensen (2007); Carson (2006); Stokke (2013)) – can be used to the advantage of accounts of this type.

First, I briefly explore the challenges posed by modelling lying (section 5.2) in a Stalnakerian common ground framework. I then argue that in both an unofficial common ground account and in the workspace account, an analysis of bald-faced lies as lies does not adequately account for the temporary acceptance of the proposition expressed by the bald-faced lie and does not account for the fact that a bald-faced lie is only successful if it 'goes on the record' (section 5.3.1). Next, I argue that an analysis of bald-faced lies as fictional statements improves upon the analysis of bald-faced lies as lies in both frameworks but that the workspace account has a definite advantage because it accounts for the aspect of 'going on the record' by analysing fictional statements as triggering parafictional updates of the form 'In/According to story s, ϕ' (section 5.3.2). Lastly, I briefly introduce and counter five possible objections to my view (section 5.4).

5.2 Common ground and lying

Just as fictional discourse poses a challenge to the traditional common ground framework that is based on cooperative information exchange (as discussed in chapters 3 and 4), modelling lying poses yet another challenge. If we follow consensus and analyse lies as a type of assertion, we can – on either an unofficial common ground or workspace account – no longer construe assertions as proposals to update a belief-based (official) common ground; obviously when Alice said (44) she did not herself believe that she had a cold (i.e., $\neg \mathbf{B}_a p$) and hence did not propose that this became common belief. Rather, after a successful deceptive lie, a correct description of the speaker's a and hearer's b beliefs would be as follows:

$$\begin{array}{cc}
 & \mathbf{B}_b p \\
\mathbf{B}_b\mathbf{B}_a p & \mathbf{B}_a\mathbf{B}_b p \\
\mathbf{B}_a\mathbf{B}_b\mathbf{B}_a p & \mathbf{B}_b\mathbf{B}_a\mathbf{B}_b p \\
\vdots & \vdots
\end{array}$$

Even though b does believe that p, and the fact that b believes that p is common belief, a does not believe that p.

I argue that there are two potential fixes to the common ground definition of assertion so that it can include lies: First, a switch to a doxastically neutral (i.e., acceptance-based) conception of the common ground, and, second, a moderate psychologistic turn (i.e., a switch to talking about the hearer's conception of common belief).

5.2.1 Acceptance-based common ground

Stokke (2013) opts for the first strategy and defines assertions as proposals to update what is commonly accepted:

$$\begin{array}{cc}
\mathbf{A}_a p & \mathbf{A}_b p \\
\mathbf{B}_b\mathbf{A}_a p & \mathbf{B}_a\mathbf{A}_b p \\
\mathbf{B}_a\mathbf{B}_b\mathbf{A}_a p & \mathbf{B}_b\mathbf{B}_a\mathbf{A}_b p \\
\vdots & \vdots
\end{array}$$

In words, a's assertion is a proposal for both a and b to accept that p, for both to believe that the other accepts that p, etc. The above definition of assertion includes deceptive lies because, by lying, Alice proposes that it becomes commonly accepted that she has a cold; Alice herself accepts this and proposes that Brian also accepts this (because he comes to believe it).

A potential problem with this move is that it obfuscates the difference between official and unofficial common grounds or between the workspace and the common ground. This problem is especially pressing for an Eckardt-style unofficial common ground account (see chapter 3) where the content of both official and unofficial common grounds is accepted persistently. If unofficial and official common grounds are all acceptance-based, there would essentially be no difference between fictional and non-fictional common grounds. In Stokke's account and the workspace account, if everything is acceptance-based, the only real difference between fictional and non-

fictional discourse is whether you accept it temporarily or persistently.[3] Also this doesn't seem to be enough to model the difference between fiction and non-fiction. As Keiser phrases it: "my carrying on a pretense with you is not a function of the length of time that I want us to accept what is being said, but rather what kind of attitude we take to what is being said" (Keiser, 2016, p.476).

5.2.2 The hearer's conception of common belief

Opting for the second strategy would entail defining assertions as proposals to update the hearer's presuppositions, i.e., the hearer's beliefs about what is common belief:[4]

$$\mathbf{B}_b p$$
$$\mathbf{B}_b \mathbf{B}_a p$$
$$\mathbf{B}_b \mathbf{B}_a \mathbf{B}_b p$$
$$\vdots$$

In words, a's assertion is a proposal to make b believe that p, believe that a believes that p, etc. The iteration implies nothing, however, about a's own beliefs. This definition of assertion includes deceptive lies because, by lying, Alice proposes that Brian believes the asserted content to be common belief. In the case of truthful assertion (where the speaker already believed the asserted content, i.e., $\mathbf{B}_a p$) such a proposal boils down to a proposal to update common belief (assuming that the speaker also believes the assertion to be successful, i.e., $\mathbf{B}_a \mathbf{B}_b p$, $\mathbf{B}_a \mathbf{B}_b \mathbf{B}_a p$, etc.).

A potential issue with this strategy is that it involves a move away from the traditional Stalnakerian framework – where speech acts are defined as proposals to update a common ground – towards a psychologistic or Gricean framework – where speech acts are defined as proposals to update an agent's mental states (i.e., the hearer's beliefs about what is common ground). In the context of this dissertation this is undesirable because our starting point was to model fictional statements and related speech acts in terms of common ground updates. The move is 'moderate' because this framework still makes use of the concept of common ground.

---

[3]Here and in the rest of this chapter I am assuming a version of Stokke's account where unofficial common grounds are truly temporary (see the discussion in section 3.4).

[4]Cf. the discussion in van Ditmarsch et al. (2020).

5.2.3 Comparing the acceptance-based and psychologistic strategies

A potential motivation to go for the acceptance-based definition of common ground is that it seems to be in line with Stalnaker's original conception of common ground. Stalnaker argues that common ground should be construed in terms of common acceptance rather than common belief. This is because sometimes people "at least temporarily, and perhaps in a limited context" (Stalnaker, 2002, p.176) ignore the possibility that some proposition is false in order to enable smooth conversation. This for instance happens when one conversational participant is aware of a defective context (i.e., notices a divergence in the conversational participants' beliefs about what is common ground) and responds by accommodating for the defect by temporarily accepting something that she believes to be false as common ground. For instance, a hearer may interpret the speaker as making a mistake; a father holding his baby girl may respond to a colleague's question "How old is he?" by saying "16 months". Stalnaker's analysis of this situation is that the hearer chooses not to address the perceived defective context but to resolve it by accepting – for the purpose of the conversation – the information "my baby is a boy'. Deceptive lies such as (44) are the mirror image of this situation; here the speaker (i.e., Alice) is the one that is in the know of the defect in context and the one that decides not to address it but to accept the asserted content (i.e., that Alice has a cold) as true to make the context non-defective.

Stalnaker offers a fair description of the mistake and liar case (i.e., I agree that the content expressed by a lie (or mistake) is at least accepted temporarily for the purpose and duration of the conversation). However, in both the workspace account and in Stokke's version of the unofficial common ground account, a distinction is made between things that are accepted only *temporarily* (for the purpose and duration of some fictional conversation) and things that are accepted more *persistently*. In the workspace account, this is the distinction between information that was temporarily accepted (as part of the workspace) also updating the stable common ground or not (only in an embedded form). In Stokke's account, this is the distinction between information being part of a temporary unofficial common ground or information being part of the stable official common ground. Stalnaker does not distinguish between temporary and stable common grounds and only establishes that the notion of common ground *in general* cannot be in terms

of common belief. However, once we introduce this distinction, it becomes clear that establishing that a lie (or mistake) involves acceptance that is temporary (and "perhaps in a limited context") of the expressed content is really only an argument for why workspaces or unofficial common grounds should be acceptance-based. It does not establish that the *stable* (official) common ground must be construed as acceptance-based. Lies and mistakes could just involve updating some kind of unofficial common ground or workspace that is only there for the duration of the conversation.[5] It is an open question whether lies or mistakes also involve updating the stable official common ground.

One way to think about this question is by trying to establish what a successful deceptive lie is (so as to establish what it is that the liar proposes). Suppose Alice was successful in her lie to Brian that she had a cold. Is it then, after the conversation has ended, stable common ground between Alice and Brian that she had a cold (because they both permanently accept this)? Or is it, after the lie, not *really* common ground that Alice had a cold but now Brian falsely believes that this is stable common ground (i.e., because the hearer *did* and the speaker *did not* update their conception of the official common ground with this information)? Or, in case someone interprets the speaker as making a mistake, is it after the conversation between the parent and the colleague stable common ground between them that the baby is a boy? Or is it not *really* common ground, but the colleague now thinks that it is (i.e., because the speaker *did* and the hearer *did not* update their conception of the official common ground with this information)? The answer to these questions will ultimately depend on how we want to use the concept of (official) common ground; whether we allow that someone can, after engaging in conversation, be aware of a stably defective context (and not do anything about it but pretend that it isn't defective in the relevant conversations), or whether we think defective contexts must always be resolved by stable acceptance, i.e., a permanent pretence on the part of the accommodator. In other words, whether we want to use the concept of common ground solely to explain linguistic behaviour or whether we also think there is a notion of 'real' common ground between two interlocutors that is dependent on their actual persistent beliefs.

---

[5]In fact, Stalnaker's comment that accommodation involves "at least temporary" acceptance hints in this direction; Alice and the father are only temporarily accepting something as common ground.

Both strategies discussed are possible in an unofficial common ground account as well as in a workspace account. Stokke opts for an acceptance-based official common ground. As will become clear below, this is a necessary move in his framework because of his analysis of bald-faced lies as lies. Since I prefer an analysis of bald-faced lies as fictional statements I am not forced towards either of the two strategies. When discussing my analysis of bald-faced lies in this chapter, I opt for a moderate psychologistic turn (although the arguments in the succeeding sections in favour of an analysis of bald-faced lies as fictional statements also work for acceptance-based variants of the theories). I opt for this strategy because it allows us to model what is interesting about the deceptive lie and how it differs from regular assertions. Namely, that successful lies create an asymmetry in the beliefs of the speaker and hearer (i.e., a defective context emerges that is not resolved).

5.3 Bald-faced lies

Now that we have an overview of the different possible accounts in place to model fiction and lying in common ground terms, I turn to bald-faced lies. I argue that an analysis of bald-faced lies as lies misdescribes the temporary acceptance of the expressed content and misdescribes the success conditions of the speech act (section 5.3.1). Next, I propose that an analysis of bald-faced lies as fictional statements improves upon an analysis of bald-faced lies as lies and can, in the workspace account specifically, adequately account for the success conditions of the speech act (section 5.3.2).

Consider another example of a bald-faced lie: Carson's (2006) cheating student scenario. A student accused of plagiarism is called to the dean's office. The student knows he plagiarized, the dean knows he did, the student knows that the dean knows, etc. However, it is also well known that the dean will not punish anyone who explicitly denies their guilt. When asked the student therefore says:

(9)   I didn't cheat on the exam.

Similarly to the Pentangeli case, the speech act cannot involve an intention to deceive the hearer (i.e., it is common belief that the student did cheat just like it was common belief that Pentangeli knew the Godfather). Yet, the student is (bald-faced) lying.

5.3.1 Bald-faced lies as lies

In this section I argue that an analysis of bald-faced lies as lies is inadequate. Stokke (2013; 2018) adopts this approach in his unofficial common ground account. This means bald-faced lies are analysed as proposals to update the official common ground.[6] A common ground updated with (45) is represented as follows:

(46)

$$
\boxed{
\begin{array}{|c|l}
\hline
x, y & \text{Offical} \\
\hline
\text{Pentangeli}(x) & \\
\text{Godfather}(y) & \\
\cdots & \\
\hline
\neg\ \boxed{\text{know}(x, y)} \\
\end{array}
}
$$

Alternatively, in the workspace account, an analysis of bald-faced lies as lies would mean they trigger assertive closure, i.e., update the stable common ground:

(47)

$$
\left\langle
\boxed{
\begin{array}{|c|l}
\hline
x, y & C \\
\hline
\text{Pentangeli}(x) & \\
\text{Godfather}(y) & \\
\cdots & \\
\hline
\end{array}
}
,
\boxed{
\begin{array}{|c|l}
\hline
x', y' & W \\
\hline
\text{Pentangeli}(x') & \\
\text{Godfather}(y') & \\
\cdots & \\
\hline
\neg\ \boxed{\text{know}(x', y')} \\
\end{array}
}
\right\rangle
$$

(48)

$$
\left\langle
\boxed{
\begin{array}{|c|l}
\hline
x', y' & C \\
\hline
\text{Pentangeli}(x') & \\
\text{Godfather}(y') & \\
\cdots & \\
\hline
\neg\ \boxed{\text{know}(x', y')} \\
\end{array}
}
, \varnothing
\right\rangle
$$

This analysis forces a switch to an acceptance-based notion of the (official) common ground (i.e., Pentangeli's speech act cannot be a proposal to update the hearer's conception of what is common belief since it is already common belief that Pentangeli knew the Godfather). Pentangeli thus does not propose that anyone *believes* that he knew no Godfather (he does not intend to deceive

---

[6]Here I obviously gloss over the fact that (45) is a fictional example.

anyone) but instead proposes that it becomes persistently commonly *accepted* that he knew no Godfather. Similarly, the student proposes that he and the dean persistently *accept* that he did not cheat (and that they both believe that the other accepts this, etc.)

Temporary acceptance

The resulting characterization of the bald-faced lie is unsatisfactory because it does not adequately model the temporary acceptance of the proposition expressed by the bald-faced lie. In the above versions of the workspace and unofficial common ground accounts, the (official) common ground is construed as involving persistent acceptance. However, bald-faced lie scenarios seem to be characterized by *temporary* acceptance.[7] If a bald-faced lie is a lie, the proposition *p* expressed by it is accepted persistently (also after the conversation has ended). This would mean that *p* would be common ground in subsequent non-fictional discourse between the same interlocutors. However, suppose that after the courtroom meeting one of the senators came up to Pentangeli in the hallway and asked him: "Hey, now that we're off the record, you did actually serve under Corleone, right?". Their conversation will continue as it did before the hearing (when it also was common ground that he knew the Godfather) and Pentangeli can be expected to answer something like: "Well, yes of course! We all know that. But I couldn't say that during the hearing!". Similarly, the dean may run into the student at a bar later that week and ask a similar question: "So, tell me, off the record, how exactly did you manage to cheat on that exam?" to which the student may reply: "Really? Off the record? Well, what I did is...". An analysis of bald-faced lies as proposals to persistently accept the expressed content would have to predict that such discourse cannot take place (or is fictional). It seems that in this sense the bald-faced lie scenarios are better described as involving a kind of temporary acceptance for the purpose and duration of the relevant conversation, i.e., while the conversation is 'being recorded' (similar to the temporary acceptance involved in workspaces or Stokke's unofficial common grounds). In other words, Pentangeli and the student are only temporarily in 'warranting contexts' (cf. Carson (2006); Fallis (2009); Saul (2012)) where they have an obligation to justify the truth of what they

---

[7]A similar point has recently been made by Keiser (forthcoming).

say when challenged.[8] Surely, in the courtroom or in the dean's office, Pentangeli and the student will need to defend what they said if challenged, but once they are off the record they can admit that what they said was false.

Going on the record

A second problem with an analysis of bald-faced lies as lies is that it misconstrues the success conditions of the speech act. The purpose of bald-faced lies is often described as that Pentangeli and other bald-faced liars want to 'go on the record' with what they said. The desire to 'go on the record' is referred to by for instance Stokke (2013), Sorensen (2007) and Carson (2006) to argue that bald-faced lies must be assertions and hence lies; the bald-faced liar wants her speech act to be 'recorded' and this supposedly only happens through assertion and not through fiction because fiction involves only temporary acceptance. However, there is little consensus on what 'going on the record' means exactly. I take some agent *a* putting some statement *p* 'on the record' as meaning that *a* ensures that it becomes stable common ground that *p* was part of a certain conversation or discourse *d*. This means that the bald-faced liar who says *p* does not actually propose a stable common ground update with *p* (as in the case of assertion), but with something of the form 'In/According to conversation *d*, *p*'. In other words, Pentangeli does not propose that anyone actually persistently accepts or believes that he knew no Godfather – and that this becomes stable common ground. This is irrelevant for Pentangeli and the Godfather.[9] Rather, what's important for the success of his bald-faced lie is that it becomes common ground that he said the right things during the hearing.[10] Specifically, he needs the Godfather (one of Pentangeli's hearers in the courtroom) to believe this. Such an update of the common ground will actually ensure that Corleone is not convicted (and hence save Pentangeli's family). Pentangeli's bald-faced lie is thus successful if something like the following hedged or modalized proposition becomes stable common ground: 'According to the courtroom proceedings, Pentangeli did not know the Godfather'. Similarly,

---

[8]See section 5.4.3 on disagreement with a bald-faced liar.

[9]Pentangeli only cares about people temporarily accepting that he knew no Godfather in so far as that what happens in the courtroom conversation determines what becomes part of proceedings. In other words, he needs people to 'play along' in court but nobody needs to persistently accept what he said as part of the (official) common ground.

[10]Cf. Leland (2015); Dynel (2011); Harris (forthcoming).

the student's speech act is successful if it becomes stable common ground between him and the dean that he did not confess but rather explicitly denied having cheated, i.e., that 'According to the conversation in the dean's office, the student did not cheat on the exam'.

5.3.2 Bald-faced lies as fictional statements

I propose, contra Stokke, to analyse bald-faced lies as fictional statements.[11] An analysis of bald-faced lies as fictional statements allows us an independent choice of strategy in dealing with lies in the common ground framework, i.e., the moderate psychologistic turn is still available.

For a psychologistic version of the unofficial common ground accounts, an analysis of bald-faced lies as fictional statements would entail that Pentangeli's speech act (being a fictional statement) is analysed as a proposal to update (or create) the hearer's conception of the unofficial common ground of the courtroom conversation (and assertions and lies as proposals to update the hearer's conception of the belief-based official common ground). Assuming it was already official common ground who Pentangeli and the Godfather were, the complete common ground updated with (45) thus looks as follows:

$$(49) \quad \left\langle \begin{array}{|c|} \hline x, y \\ \hline \mathrm{Pentangeli}(x) \\ \mathrm{Godfather}(y) \\ \hline \end{array} \text{Official} \quad , \quad \begin{array}{|c|} \hline x, y \\ \hline \mathrm{Pentangeli}(x) \\ \mathrm{Godfather}(y) \\ \hline \neg \boxed{\mathrm{know}(x, y)} \\ \hline \end{array} \text{Unofficial}^{court} \right\rangle$$

In a psychologistic version of the workspace account an analysis of bald-faced lies as fictional statements entails that, whereas assertions and lies trigger assertive closure, fictional statements and bald-faced lies trigger fictive closure. In other words, the content of the hearer's conception of the workspace (e.g., that Frodo had a very trying time that afternoon or that Pentangeli knew no Godfather) is not added directly to the hearer's

---

[11]This understanding of bald-faced lies comes closest to Maitra's (2018) who discusses similarities between bald-faced lies and the utterances of an actor on stage. However, Maitra has a different understanding of 'going on the record' that is deemed irrelevant to what the actor and bald-faced liar do. Moreover, I prefer to stick with the term 'fictional statement' so as to not exclude the possibility of written bald-faced lies.

conception of what is common belief. Rather, it is added to the hearer's conception of what is common belief as modalized, parafictional information, i.e., respectively under the 'In/According to *The Lord of the Rings*'-operator and the 'In/According to the courtroom proceedings'-operator ($\square_{court}$):

(50) $\left\langle \begin{array}{|c|}\hline x, y \quad \text{C} \\ \hline \text{Pentangeli}(x) \\ \text{Godfather}(y) \\ \ldots \\ \hline\end{array} \; , \; \begin{array}{|c|}\hline x', y' \quad \text{W} \\ \hline \text{Pentangeli}(x') \\ \text{Godfather}(y') \\ \ldots \\ \neg\; \boxed{\text{know}(x', y')} \\ \hline\end{array} \right\rangle$

(51) $\left\langle \begin{array}{|c|}\hline x, y \quad \text{C} \\ \hline \text{Pentangeli}(x) \\ \text{Godfather}(y) \\ \ldots \\ \square_{court}\; \begin{array}{|c|}\hline x', y' \\ \hline \text{Pentangeli}(x') \\ \text{Godfather}(y') \\ \ldots \\ \neg\; \boxed{\text{know}(x', y')} \\ \hline\end{array} \\ \hline\end{array} \; , \varnothing \right\rangle$
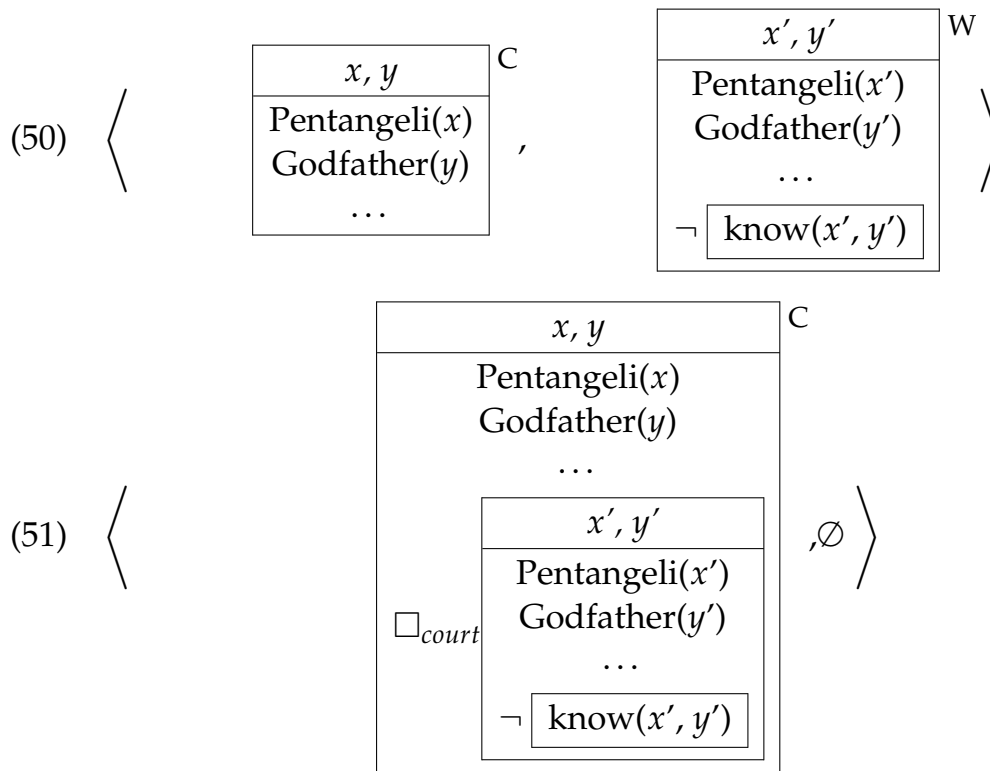
Temporary acceptance

An advantage of an analysis of bald-faced lies as fictional statements is that it enables us to make sense of the fact that the propositions expressed by bald-faced lies are only accepted temporarily.[12] If we adopt an analysis of bald-faced lies as fictional statements in Stokke's unofficial common ground account, bald-faced lies and fictional statements update a *temporary* unofficial common ground, i.e., the interlocutors temporarily accept that Pentangeli knew no Godfather or that Frodo had a trying time some afternoon. At the end of the discourse – as the hearing ends or as we stop reading *The Lord of the Rings* – the unofficial common ground evaporates. Hence the content expressed by the bald-faced lie is only accepted temporarily.

---

[12]On an Eckardt-style version of the account (see chapter 3), unofficial common grounds are non-temporary so this aspect of bald-faced lies remains unaccounted for.

In the workspace account, bald-faced lies and fictional statements update a *temporary* workspace during the discourse. At the end of the discourse, fictive closure is triggered and the expressed propositions – that Pentangeli knew no Godfather or that Frodo had a trying time some afternoon – are no longer accepted in their unembedded form (but a prefixed version *is* added to the hearer's conception of the stable common ground). Hence, as in the unofficial common ground account, the content expressed by the bald-faced lie is only accepted temporarily.

Going on the record

Another advantage of the analysis of bald-faced lies as fictional statements is that in the workspace account it allows us to account for the fact that bald-faced lies are successful if they 'go on the record'. I have argued that Pentangeli's and the student's bald-faced lies are proposals to make it stable common ground that 'According to the courtroom proceedings, he knew no Godfather' and 'According to the conversation in the dean's office, the student did not cheat on the exam'. Crucially, these are *parafictional* statements of the form 'In/According to story s, ϕ' (i.e., $\Box_s \phi$). In the workspace account, a bald-faced lie p (or a fictional statement p) is analysed as ultimately a proposal to make the hearer's conception of what is common belief include $\Box_s p$. Hence the workspace account can model how the bald-faced liar succeeds at 'going on the record' with their speech act, i.e., making $\Box_s p$ stable common ground (assuming that the bald-faced liar also believes that their speech act was successful). In the unofficial common ground accounts there are no parafictional updates of the official common ground at the end of fictional discourse and hence this aspect of bald-faced lies remains unaccounted for.[13]

5.4 Possible objections

Now that I have made my case why, in terms of common ground updates, the bald-faced lie is much more akin to a fictional statement than to a lie, I will introduce five possible objections to this analysis and offer replies.

---

[13]As discussed in chapter 3, we could model parafictional information as 'stored' in the relevant unofficial common ground but this forces us to give up that unofficial common grounds are temporary.

5.4.1 Non-fictional parafictional updates

A possible objection to the analysis outlined above is that updates of the form 'According to the courtroom proceedings, he knew no Godfather' are also possible in non-fictional discourse. For instance, Stalnaker (2002) maintains that, since speech acts are so-called 'manifest events', as a side effect of a asserting p, it will also become common ground that a said p in some conversation or discourse d. Arguably, if it is also common ground that a's assertion was accepted, this will imply that it was common ground in d that p (or $\Box_d p$). So can 'going on the record' not also be achieved through regular assertions or lies?

Although I agree that people can also 'go on the record' with non-fictional discourse, this does not establish that bald-faced lies could just as well be analysed as regular assertions. The crucial difference between fictional statements and assertions is that this 'parafictional update' in the case of non-fiction is a *non-essential* update (as it is for Stalnaker) whereas it is an *essential* update in the case of fiction. In other words, when a asserts p, a proposes to make p common ground (through the workspace), not to make p and $\Box_d p$ common ground. On the other hand, when a fictionally states p, a proposes to make $\Box_d p$ common ground (through the workspace). Hence an assertion is also successful if the parafictional update does not take place but a fictional statement is not (e.g., Tolkien's fictional statement (25) is not successful if, after reading *The Lord of the Rings*, it is not common ground that in *The Lord of the Rings*, Frodo had a very trying time). This is why bald-faced lies are like fictional statements; a bald-faced lie p is only successful if we temporarily accept p and as a result of that the common ground is updated with $\Box_d p$.

Moreover, even if we would (contra Stalnaker) construe assertions as involving a parafictional update as part of their essential updates (i.e., define assertions as proposals to update the common ground with p and $\Box_d p$ and fictional statements as proposals to update the common ground with $\Box_d p$), bald-faced lies would *still* be more like fictional statements precisely because (like fictional statements) their success does not depend on whether the common ground is updated with the content they express but only on whether it is updated with the relevant parafictional update.

5.4.2 Commitments in bald-faced lie scenarios

Another possible objection to the current proposal is that conversations in institutional settings, such as in a courtroom or in the dean's office, seem to be eminent models of conversations where what people commit themselves to (i.e., what 'goes on the record') is essential for determining subsequent actions and sanctions in the real world. Bald-faced lies such as Pentangeli's and the student's take place in such contexts and hence it is counterintuitive to analyse them as akin to fictional statements. Fictional statements do not impose commitments on speakers like assertions do, so how can they determine possible sanctions? In response to this, I suggest that in fact the whole conversation in a bald-faced lie scenario can be understood as a pretend-conversation. To make this intuitive, suppose that Pentangeli *had* in fact wanted to harm the Godfather and testified against him in court (or that the student, for some reason, wanted to be punished and confessed in the dean's office):

   (52)   Yes, I knew the Godfather.

In this scenario it still does not matter to Pentangeli whether anyone somehow comes to accept or believe that he knew the Godfather. Actually, everyone already believes this and Pentangeli is aware of this. So, again, his utterance of p is not aimed at making p persistent common ground.[14] To really hurt the Godfather, Pentangeli needs it to become common ground that 'According to the courtroom proceedings, Pentangeli knew the Godfather'. Again, the speech act is aimed at a parafictional update that results from fictive closure:

---

[14]Harris (forthcoming) makes the similar point that a witness who already admitted everything on the courthouse steps, with everyone present who is also in the courtroom, still will have to repeat this confession in the courtroom in order to affect subsequent legal sanctions.

$$(53) \quad \left\langle \quad \boxed{\begin{array}{c|c} \multicolumn{2}{c}{x, y} \quad \text{C} \\ \hline \text{Pentangeli}(x) \\ \text{Godfather}(y) \\ \cdots \\ \square_{court} \boxed{\begin{array}{c} x', y' \\ \hline \text{Pentangeli}(x') \\ \text{Godfather}(y') \\ \cdots \\ \text{know}(x', y') \end{array}} \end{array}} \quad ,\varnothing \quad \right\rangle$$
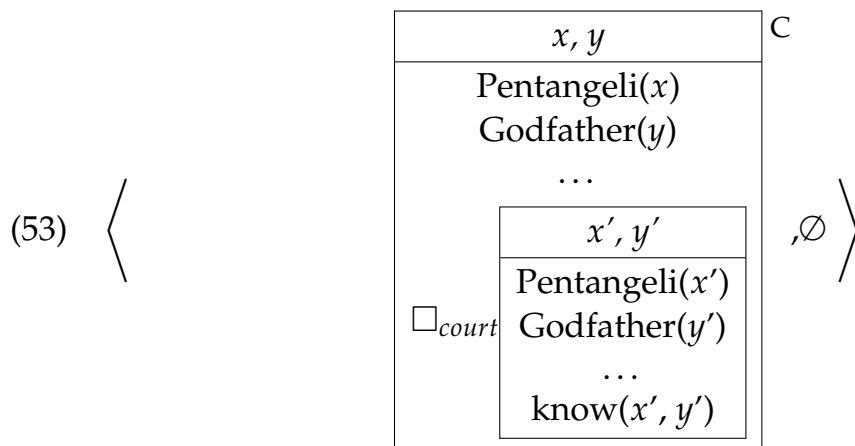
So not only bald-faced lies but also other utterances that could be made in bald-faced lie scenarios (e.g., (52)) should be understood as fictional statements.[15]

Likewise, the whole conversation in the courtroom or in the dean's office can be understood as a kind of pretend-conversation, i.e., as consisting of fictional statements.[16] Even if everyone present in court already knows that someone is in fact guilty, still the parties involved will speak as if they do not yet know this and are trying to establish what happened based on the available testimonies and evidence. As Keiser (2016) helpfully suggests, the courtroom conversation is much like a proof; you may not assume its conclusion even if you already know it to be true. What determines potential sanctions is what we eventually have been able to 'prove' or establish as 'true according to the courtroom conversation'. For instance, even though it may be common ground that someone is guilty, if we cannot 'prove' this in an institutional setting, they are considered legally innocent. Hence, even though the conversations in court or in the dean's office are pretend-conversations, the commitments made *within* these conversations are pivotal for determining subsequent action; the way the pretend-conversation progresses will determine the relevant parafictional data and this data in turn has 'real world' legal consequences.

---

[15]Not *all* possible utterances in bald-faced lie scenarios constitute continuations of the shared pretence. See footnote 18.

[16]In this respect the analysis is similar to Keiser's (2016) account according to which bald-faced lies are moves in language games; Pentangeli makes a move in the courtroom game by which he avoids scoring points against the Godfather. In the workspace account these language games are understood as pretend-conversations and 'scoring points' is analysed as ensuring that the desired parafictional updates take place.

Additionally, this analysis explains why Pentangeli and the student are being asked questions in the first place that are somehow 'insincere' because the person asking already knows the answer to them; in the pretend-conversation the questions asked *are* sincere because it is part of the pretence that the questioner does not yet know the answer to them. Moreover, the analysis accounts for a special feature of courtroom talk of being able to 'scratch' or undo statements. A similar 'scratching' is possible in pretend play where you can stop and rewind pretend scenarios to revise what will eventually be true in the fiction.[17] By contrast, taking back something you have asserted requires retraction of something that was said; you cannot simply replay the discourse and make it true that you never said it.

5.4.3 Disagreement with a bald-faced liar

In line with the previous concern, Marques (2020) has recently argued that the fact that we can (and often do) disagree with bald-faced liars shows that bald-faced lies cannot be fictional statements. People contradict bald-faced lies because it is important for them to 'set the record straight'. We don't feel compelled to (and shouldn't) contradict fictional discourse. For instance, you wouldn't interrupt an actor on stage because you believe that they uttered a falsity.

The analysis of bald-faced lie conversations as pretend-conversations explains why people can (and often will) disagree with a bald-faced liar in institutional settings. The disagreement takes place *within* the pretend-conversation and is crucial for determining what eventually is parafictional truth. Although I agree with Marques that we shouldn't disagree with an actor on a stage, this is because, as an audience member, you cannot influence fictional truth. However, scripted performances are not the only kind of fictional discourse. Courtroom conversations are probably more akin to unscripted pretend-discourse (e.g., 'improv' or role-playing games) which *can* involve disagreement. Consider the following pretend-conversation:

---

[17]This seems to be what the senator questioning Pentangeli initially attempts in reply to the bald-faced lie: "Mr. Pentangeli, you – you are contradicting a sworn statement you've previously made to me and signed. I ask you again, sir, here and now under oath: Were you at any time a member of a crime organization headed by Michael Corleone?"

(54) *Tineke*: Help! A dragon is approaching. Flee!
      John: Don't worry, it is actually a unicorn.
      Tineke: Eh, no... It really is a dragon and now it's spitting fire on you!

Here two 'actors' engaged in a pretend-conversation are both equally licensed to determine what is fictionally common ground and hence what is true in their game of pretence. There is no *real* disagreement between Tineke and John whether there is a dragon approaching (it is stable common ground that dragons do not exist) but there is a temporary pretend-disagreement. Their pretend-disagreement over what kind of creature they are facing influences whether, after the pretend play, the parafictional information that "In the pretend play, a dragon approached John and Tineke" is stable common ground or not.

   Similarly, the senator questioning Pentangeli or the dean questioning the student can disagree with the bald-faced liar and push their interlocutor on what they said *within* the pretend-discourse. Although the questioners themselves would probably not be allowed to respond with "That's not true/a lie",[18] they *can* report on someone having directly contradicted the bald faced liar. The bald-faced liar, or another party in the pretend-conversation that wants the content expressed by the bald-faced lie to become parafictional truth, will then respond by defending the bald-faced lie. Hence we can expect, as Marques also notes, that bald-faced liars will often accuse the party that contradicts them of being the liars. For instance, the courtroom discourse in *The Godfather II* continues as follows:

(55) *Senator*: We have a sworn affidavit – we have it – your sworn affidavit that you murdered on the orders of Michael Corleone. [...]
      Pentangeli: Look, the FBI guys, they promised me a deal. So, so I

---

[18] Since it is part of the institutionalized pretend play that the conversation is a cooperative information exchange where the questioner is trying to find out what happened, the dean's or senator's utterance of "That's a lie/false" would be inadmissible. In these contexts such a statement would not serve its ordinary function of signalling that an attempt to deceive has failed. Rather, it functions like a notification that the questioner is refusing to play along with the bald-faced liar and terminates the pretend-discourse. In some bald-faced lie scenarios (e.g., the cheating husband or chicken thief scenarios described in section 5.4.4) such a refusal would be an admissible move but in institutional settings, officials such as the dean and the senator are not allowed to 'break character' in this way. Arguably, another witness could potentially accuse the bald-faced liar of lying without breaking character in the pretend scenario).

> made up a lot of stuff about Michael Corleone 'cause that's what they wanted. But – But it was all lies. Uh... everything!

Again, there is no *real* disagreement between the senator and Pentangeli over whether Pentangeli knew the Godfather; it is stable common ground that he did. There is, however, a temporary pretend-disagreement between the interlocutors, through which both the senator and Pentangeli attempt to fix the record (i.e., influence the parafictional information that eventually becomes stable common ground).

5.4.4 The point of fiction and bald-faced lies

An additional possible concern with my view is that bald-faced lies and fictional statements seem to differ essentially in the point of the speech acts. Both Pentangeli's and Tolkien's speech acts can be construed as proposals to update the workspace and trigger fictive closure, but only for Pentangeli does the parafictional update really seem to be the point of the speech act; the point of Tolkien's *The Lord of the Rings* is to make its readers imagine its content (which involves updating the workspace) and enjoy that experience, not to stock up on parafictional beliefs.

First, it is important to note that it is unclear whether a difference in the point of bald-faced lies and fictional statements is relevant for deciding what kind of speech acts bald-faced lies are; *why* people are performing some speech act (e.g., telling a fictional story) seems to be an essentially different question from *whether* people are performing this speech act. Still, it is interesting to see whether and why the motivation behind performing these two types of speech acts may diverge.

Second, it is not obvious that this difference in motivation generalizes to all stereotypical fictional statements. Suppose the CEO of a firm calls in an employee, Mike, after Mike disagreed with him during a meeting. The CEO tells the following story:

(56)   Listen, Mike. Once upon a time there was an annoying little midget that pissed on a giant's cake. Do you want to know what the giant did to the midget? He ate him, Mike.

Is the point of this fictional story to make Mike temporarily imagine a story about giants and midgets or is the point of this story to make Mike believe that 'In the CEO's story, the giant eats the midget that bothered him' and draw the 'appropriate' lessons from that?

Third, even if the point of fiction is generally to make its audience entertain its content, sometimes this is the point of a bald-faced lie as well. Pentangeli's statement in court and the student's statement in the dean's office are examples of bald-faced lies in rule-bound institutional settings. However, bald-faced lies can also be more spontaneous. Meibauer (2014) discusses the case of a cheating husband who tells his wife that he has spent the night in his office while it is common knowledge that he was with his lover. Or consider Cargile's (1970) bald-faced lie scenario in which a man catches his son-in-law in his chicken coop at 5 AM holding two dead chickens. The son in law then says:

(57)   My coming by must have frightened the thief away.

Arguably, the aim of these bald-faced lies is to make the people involved entertain or uphold a favourable narrative about their marriage or family relations (not to obtain parafictional beliefs about such a narrative). Such bald-faced lies are thus in this sense even more like stereotypical pretend-discourse than Pentangeli's speech act. This suggests that, if there is a difference in the point of bald-faced lies and fictional statements, it is a gradual difference where bald-faced lies and fictional statements lie on a continuum; they are similar in the sense that both result in parafictional updates but bald-faced lies can be more or less like stereotypical fictional statements depending on whether the point of the speech act is also to obtain this result.

5.4.5 Blameworthiness

The last potential worry with the analysis outlined above is that bald-faced lies and fictional statements seem to differ essentially with respect to blameworthiness; whereas the bald-faced liar is blameworthy for saying something false (e.g., Pentangeli can be convicted of perjury), someone who tells a stereotypical fictional story (e.g., Tolkien) is not blameworthy – rather, they are praiseworthy – for stating something false. If Pentangeli is just telling a fictional story, why should he be any more blameworthy than Tolkien?

First, although we can deem some bald-faced liars (e.g., the student) blameworthy (contra Sorensen (2007)), blameworthiness does not extend to all bald-faced lie scenarios.[19] Can we really blame the couple in the cheating

---

[19]It is not even clear that blameworthiness extends to all lies. Is the white liar blameworthy? Or the liar that saves innocent people with their lie?

husband scenario or the father and son-in-law in the chicken thief scenario for pretending their relation is better than it is? Moreover, many examples of bald-faced lies in institutional settings (e.g., the case of Pentangeli) involve some kind of coercion into making a particular statement. The Godfather was non-verbally threatening to hurt Pentangeli's family if he would testify against him. Can we really blame Pentangeli for working the system to protect his family?

Second, in cases where the bald-faced liar does seem to be blameworthy (to some extent) I follow Maitra (2018) who argues that this is because of an implicit or explicit 'honor code' that requires the bald-faced liar to not just 'tell some story' but be truthful in the relevant context. The parafictional updates that result from their speech acts should reflect what we know is really the case.[20] For instance, the student is blameworthy for (by telling a fictional story) creating an unfair academic advantage for himself compared to the other students. Likewise, Pentangeli is to some extent blameworthy for creating an unfair trial. Such honor codes are not in place when people tell stereotypical fictional stories or in the described cases of spontaneous bald-faced lying. Moreover, given that there is an implicit or explicit expectation for the courtroom (or conversation with the dean) proceedings to track the truth, bald-faced liars can also be construed as being blameworthy for (indirectly) deceiving an potential third party that is not aware of the falsity of the bald-faced liar's statement and would read the proceedings expecting them to be truthful.

5.5 Conclusions

In this chapter I have explicated different possible strategies to model lies in Stalnaker's common ground framework; we can either switch to an acceptance-based conception of the stable (official) common ground or we can take a moderate psychologistic turn and define speech acts as proposals to update the hearer's belief-based presuppositions. I have argued that an analysis of bald-faced lies as fictional statements is superior to an analysis of bald-faced lies as lies. The analysis of bald-faced lies as fictional statements

---

[20]In other words, certain content expressed in the courtroom conversation must be viable for export (see section 4.6.2). Similarly, authors of fiction can be blameworthy if they intentionally include falsities in a novel that is expected to be realistic with respect to a certain cluster of facts.

allows us to account for the temporary acceptance of the content expressed by the bald-faced lie. In a workspace account it also allows us to adequately model the success condition of the bald-faced lie of 'going on the record' as consisting in a parafictional update. In response to possible objections I have argued that even if non-fictional statements also involve a 'parafictional' update, this does not constitute an essential update of the speech act as it does in the case of bald-faced lies and fictional statements. Second, the analysis of bald-faced lies as fictional statements can be extended to the entire conversation in bald-faced lie scenarios. This allows us to explain (amongst other things) why, even when p is common ground, testifying p is just as relevant as testifying $\neg p$ and why people can disagree with bald-faced liars. Last, I have aimed to defuse some worries about supposedly essential differences between bald-faced lies and fictional statements such as the point of the speech act and blameworthiness.

6 The challenge of metafictional anaphora

This chapter is a rewritten version of 'The Challenge of Metafictional Anaphora' in *At the Intersection of Language, Logic, and Information* (itself a rewritten version of the paper 'Metafictional anaphora: A comparison of different accounts' in *Proceedings of the 2018 ESSLLI Student Session*). Parts of section 6.4.3 are adapted from 'Revisiting the 'wrong kind of object' problem' which is a co-authored paper with Prof. Dr. Edward N. Zalta in *Organon F*. The most substantial differences between this chapter and the volume contribution are: First, a removal of the introduction to the basics of the workspace account. Second, the addition of the suggestion to design more uniform fictive closure* and fictive closure** operations by combining them with a de re version of the workspace account (see especially 6.4.3). Third, the incorporation of insights from the co-authored paper with Prof. Dr. Edward N. Zalta 'Revisiting the 'wrong kind of object' problem', i.e., the specification of the desideratum posed by mixed discourse (see section 6.2.2) and details on how object theory may deal with the seeming reoccurrence of the problem of the wrong kind of object for explicit parafictional statements (see section 6.4.3).

6.1 Introduction

Up until this point in developing the workspace account (as a framework to model regular assertions, lies, fictional statements and parafictional discourse) I have assumed a simple 'descriptivist' analysis of parafictional updates (see section 4.6.1). Consider fictional statement (25) and parafictional statement (58):

(25)   Frodo had a very trying time that afternoon.

(58)   In *The Lord of the Rings*, Frodo was born in the Shire.

When engaging with a fictional statements such as (25), the workspace is (temporarily) updated with the content that there is an entity called 'Frodo' and that this entity had a trying time on some afternoon. At fictive closure, this entire workspace (including discourse referents and conditions) is copied into the common ground under a *The Lord of the Rings* fiction operator. Similarly, a parafictional statement such as (58) updates the common ground with the modalized information that in *The Lord of the Rings*, there is an entity named 'Frodo' who was born in the Shire. On the basic descriptivist version of the workspace account a common ground updated with (25) and (58) thus looks as follows:

(59) $\left\langle \quad \begin{array}{|l|} \hline \qquad\qquad\qquad\qquad\qquad\qquad\quad C \\ \hline \cdots \\ \square_{lotr} \begin{array}{|c|} \hline x \\ \hline \cdots \\ \text{Frodo}(x) \\ \text{trying.time}(x) \\ \text{born.in.Shire}(x) \\ \hline \end{array} \\ \hline \end{array} , \varnothing \right\rangle$
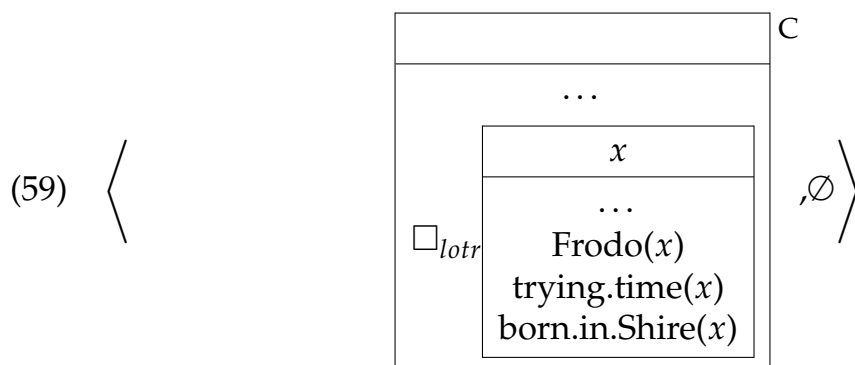
The current chapter introduces a new type of discourse that will complicate this basic picture: 'metafictional discourse'. Metafictional statements are statements about fictional entities *as fictional entities*. Consider for instance the following statements:

(6)   Frodo was invented by Tolkien.

(60)   Frodo is a fictional character.

(61)   Elijah Wood has portrayed Frodo.

Sentences (6), (60) and (61) tell us something about Frodo, not as a flesh and blood individual, but as a fictional character. Although they feature fictional names, metafictional statements are a type of assertion; (6), (60) and (61) are not just fictionally true but really true. Moreover, as Lewis (1978) argued, they cannot be reduced to implicit parafictional discourse by analysing them as covertly prefixed by a fiction operator; it is not true that in *The Lord of the Rings*, Frodo is a fictional character that was invented by Tolkien and that has been portrayed by Elijah Wood. Rather, the above statements seem to be simply true statements about some abstract object called 'Frodo'.

In this chapter I first specify the exact challenge posed by metafictional discourse using a recent insight due to Semeijn and Zalta (2021) (section 6.2). We have argued that the possibility of anaphoric dependencies and co-predication across different types of statements featuring fictional names establishes a need for a uniform semantic treatment of fictional names across metafictional and parafictional discourse (but not fictional discourse) that avoids the so-called 'problem of the wrong kind of object' (e.g., we want to avoid an analysis that implies that a flesh and blood individual was invented by someone or that an abstract object was born in a certain region). Second, I argue that in answering this challenge the workspace account runs into an accessibility problem that is best illustrated by looking at pronominal anaphora across mixed parafictional/metafictional discourse (henceforth 'metafictional anaphora') (section 6.3). This issue generalizes to other current dynamic semantic approaches to fiction such as the unofficial common ground accounts (section 6.3.2). I explore and evaluate four different possible solutions based on a descriptivist analysis of pronouns (Section 6.4.1), Maier's (2017) psychologistic DRT (Section 6.4.2), Zalta's (1983, 1988) theory of abstract objects (Section 6.4.3) and Recanati's (2018) dot-object analysis of fictional characters (Section 6.4.4).

6.2 The desideratum posed by metafictional discourse

6.2.1 The problem of the wrong kind of object

The basic puzzle that is posed by metafictional discourse is to provide a coherent semantic analysis of fictional, parafictional and metafictional discourse that avoids the problem of the wrong kind of object.[1] Prima facie, we would like to remain 'semantically innocent' and assume that fictional names such as 'Frodo' refer uniformly to the same thing in different contexts. Such a uniformity approach can take two basic forms; an 'anti-realist' approach and a 'realist' approach. Both run into variations of the wrong kind of object problem.

---

[1]This term was originally coined by Klauk (2014) for what I call the realist variant of the problem of the wrong kind of object.

First, we may adopt an anti-realist analysis of fictional names (e.g., Lewis (1978); Walton (1990); Maier (2017)) and assume that the name 'Frodo' refers uniformly to a flesh and blood hobbit. Obviously, this hobbit does not exists in the real world but only exists in the *The Lord of the Rings* worlds. Hence we call this the 'anti-realist' approach to fictional names. However, such an approach runs into difficulties with metafictional statements. Surely, a flesh and blood hobbit (that had a trying time and was born in the Shire) is not the right kind of object to also be a fictional character (that was invented by Tolkien and portrayed by Elijah Wood). This is the anti-realist variant of the problem of the wrong kind of object.

Alternatively, we may adopt a realist analysis of fictional names (e.g., Zalta (1983, 1988); van Inwagen (1977)) and assume that the name 'Frodo' refers uniformly to an abstract object. Insofar as abstract objects exist, they exist in the real world. Hence we call this approach the 'realist' approach: fictional names (e.g., 'Frodo') refer to things that actually exist (i.e., an abstract object). A realist approach to fictional names runs into difficulties with the interpretation of fictional statements such as (25) and parafictional statements such as (58). Surely, abstract objects (e.g., fictional characters) are not the right kind of objects to be having trying times or to be born in the Shire. This is the realist variant of the problem of the wrong kind of object.

6.2.2 Mixed discourse

A prima facie attractive response to the problem of the wrong kind of object is to give up semantic innocence (e.g., Kripke (2011); Currie (1990)). Apparently fictional names, unlike regular proper names, are ambiguous. A name such as 'Frodo' refers to a flesh and blood hobbit in fictional and parafictional statements, and to an abstract object in metafictional statements. This approach avoids both variants of the problem of the wrong kind of object.

Although we will see that there is a grain of truth in an analysis that posits some kind of ambiguity in fictional names, the simple approach sketched above will not do. A widely discussed problem with this strategy is the omnipresence and naturalness of anaphoric dependencies and co-predication across discourse that mixes parafictional and metafictional statements. For instance, consider the following statements:

(62) Bond is a killer but remains as popular as ever. (Collins, 2019, p.1)

(63) Morris Zapp, David Lodge's most colorful character, is a reader-response theorist modeled on Stanley Fish. (Everett, 2013, p.166)

(64) Sherlock Holmes$_i$ is a fictional character created by Conan Doyle. In Conan Doyle's stories, he$_i$ is a private detective who investigates cases for a variety of clients, including Scotland Yard. (Adapted from Recanati, 2018, p.37)

Sentences (62) and (63) are examples of co-predication of metafictional properties and (implicit) parafictional properties that apply to the same fictional name. Sentence (63) expresses that Morris Zapp is a reader-response theorist (implicit parafictional information) but also that Moris Zapp is modeled on Stanley Fish (metafictional information). Similarly, (62) expresses that Bond is a killer (implicit parafictional information) and that he remains as popular as ever (metafictional information: Bond does not remain "as popular as ever" in the fiction). Prima facie, the fact that a metafictional and an implicit parafictional predicate can simultaneously predicate over the same fictional name suggests that names such as 'Morris Zapp' cannot be ambiguous (i.e., refer to different things in parafictional and metafictional discourse). In a similar vein, (64) is an example of anaphoric dependency across a metafictional and subsequent explicit parafictional statement. The parafictional statement contains a pronoun 'he' that is anaphorically dependent on the name 'Frodo' introduced in the preceding metafictional statement. Standardly, we take this to mean that the two terms (i.e., 'Sherlock Holmes' in the metafictional statement and 'he' in the parafictional statement) co-refer. Again we have a prima facie reason to forgo an analysis that posits an ambiguity between fictional terms used in metafictional as oppposed to parafictional contexts.

Based on the perceived admissability and naturalness of co-predication and anaphoric dependencies in such mixed discourses, we might conclude that there is a strong argument for general semantic innocence, i.e., the debate on mixed discourse establishes the need for a uniform semantic treatment of fictional names across the different statements that may feature them. However, as Semeijn and Zalta (2021) have argued, this conclusion may be too quick. In fact, the existing literature only offers examples of co-predication and anaphoric dependencies across (implicit or explicit) parafictional and metafictional discourse. Attempts at anaphoric dependencies

across parafictional and fictional, or metafictional and fictional discourse are decidedly less natural. Consider the following examples:[2]

(65) In the story I made up yesterday, a wizard called Brian$_i$ falls in love with a cauldron. Let me tell it to you: One day, he$_i$ was alone in his$_i$ study trying out a new love-potion recipe... (Semeijn and Zalta, 2021, p. 6)

(66) In order to capture the witch, Mary$_i$ travelled to the woods and disguised herself as a potato.* In the woods she$_i$ encountered many perils...
*I know this is weird but I invented her$_i$ while eating chips. (Semeijn and Zalta, 2021, p. 6-7)

Although (65) and (66) are interpretable, they sound very awkward. In order to construe example discourses that clearly mix *fictional* discourse with parafictional or metafictional discourse (rather than implicit parafictional discourse with explicit parafictional or metafictional discourse), the beginning or interruption of the fictional discourse has to be clearly marked in the language (e.g., By "Let me tell it to you" in (65) or by inserting a footnote in (66)). This is why it is extremely challenging to come up with examples of co-predication across mixed fictional and meta- or parafictional discourse. Moreover, these clearly marked 'boundaries' of the fictional discourse make anaphoric dependencies across them as in (65) and (66) sound very unnatural. (65) would become more natural if we repeated the name 'Brian' at the start of the fictional discourse (or even better "Once upon a time there was a wizard named Brian who..."). Similarly, (66) would become more natural if we repeated the name 'Mary' in the metafictional footnote (or even better "this character").

In other words, although the literature on mixed discourse does establish a desideratum that the semantic treatment of fictional names across parafictional and metafictional discourse is uniform, an ambiguity analysis with respect to fictional names in fictional discourse (versus fictional names in parafictional and metafictional discourse) is still a viable option. The act of story-telling seems to be importantly distinct from parafictional and metafictional discourse.

---

[2]See the appendix for (attempts at) anaphoric dependencies across all possible types of mixed discourse with fictional, parafictional and metafictional statements.

6.3 Metafictional anaphora and accessibility

The current challenge for semanticists of fiction is thus to offer a uniform semantic account of parafictional and metafictional statements (so as to account for the admissibility of co-predication and anaphoric dependencies across parafictional and metafictional mixed discourse rather than only being able to account for these statements in isolation) that avoids the problem of the wrong kind of object in its different variants. In the dynamic framework of DRT this desideratum entails that we require parafictional and metafictional discourse to update on the same discourse referents. It seems that we are back at the two basic strategies described in section 6.2.1: assuming an analysis of parafictional discourse as embedded assertions (see chapter 4), we can adopt a realist approach where discourse referents for fictional entities are available in the main DRS (e.g., It is common ground that there is an abstract object named Frodo) or an anti-realist (descriptivist) approach where these discourse referents are embedded under the relevant fiction operator (e.g., It is common ground that in *The Lord of the Rings* there is flesh and blood hobbit named Frodo):

Realist approach   *Anti-realist (descriptivist) approach*
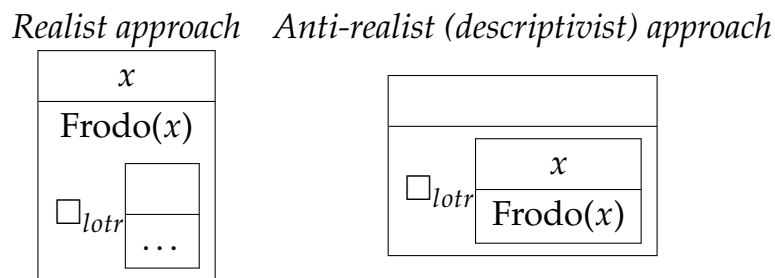
In the current descriptivist version of the workspace account, parafictional discourse operates on discourse referents that are embedded under fiction operators. Hence the workspace account is anti-realist: Frodo only exists in the *The Lord of the Rings* worlds where he is a flesh and blood hobbit. In this section I will show that this version of the workspace account runs into a specific technical problem of accessibility when dealing with metafictional discourse.

Here it is helpful to consider that the current challenge to give correct interpretations of all admissible forms of co-predication and anaphoric dependencies across parafictional and metafictional discourse subdivides into three challenges: First, account for discourse where a fictional name is introduced in a parafictional statement and a metafictional statement

contains a pronoun anaphorically dependent on that, i.e., 'metafictional anaphora' (e.g., (67)); second, account for discourse where a fictional name is introduced in a metafictional statement and a parafictional statement contains a pronoun anaphorically dependent on that, i.e., 'parafictional anaphora' (e.g., (68)); and third, account for cases of co-predication involving both parafictional and metafictional predicates (e.g., (69)):

(67) In *The Lord of the Rings*, Frodo$_i$ goes through an immense mental struggle to save his$_i$ friends. Ah yes, he$_i$ is an intriguing fictional character!

(68) Frodo$_i$ is an intriguing fictional character. In *The Lord of the Rings*, he$_i$ goes through an immense mental struggle to save his$_i$ friends.

(69) Frodo is an intriguing fictional character that goes through an immense mental struggle to save his friends.

In the following I mainly focus on the challenge posed by metafictional anaphora (Section 6.3.1) because this type of discourse gives rise to accessibility issues in the workspace account. Before turning to possible solutions in section 6.4, I will also show how accessibility issues generalize to other anti-realist dynamic approaches that also embed discourse referents for fictional entities (Section 6.3.2).

6.3.1 Metafictional anaphora in the workspace account

Consider how our central example of metafictional anaphora (67) updates the common ground in the workspace account. Both sentences in (67) (a parafictional statement about *The Lord of the Rings* and a metafictional statement about Frodo) are assertions and hence trigger assertive closure on a workspace updated with (67). First, we open up a workspace and update it with the parafictional statement in (67):

(70) $\left\langle \quad \boxed{\begin{array}{c} \\ \hline \cdots \end{array}}^{\,C} \;,\; \boxed{\begin{array}{c} \hfill W \\ \hline \Box_{lotr}\; \boxed{\begin{array}{c} x \\ \hline Frodo(x) \\ struggle(x) \end{array}} \end{array}} \quad \right\rangle$
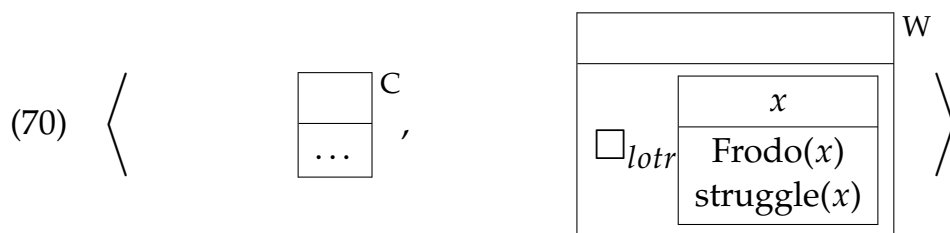
Next, we update the workspace with the metafictional statement in (67). As noted in the introduction, metafictional statements are not covertly embedded under 'In fiction *s*'-operators. In other words, there is no implicit 'In

fiction s'-operator in the second half of (67), so the workspace is updated as follows:

$$
(71) \quad \left\langle\ \boxed{\begin{array}{c} \\ \hline \cdots \end{array}}\ \text{C} \ ,\ \boxed{\begin{array}{l} \hspace{5em} \text{W} \\ \hline \Box_{lotr}\ \boxed{\begin{array}{c} x \\ \hline \text{Frodo}(x) \\ \text{struggle}(x) \end{array}} \\[1.5em] \text{intriguing}(y) \\ \ulcorner\! \boxed{\begin{array}{c} y \\ \hline he(y) \end{array}}\!\urcorner \end{array}}\ \right\rangle
$$
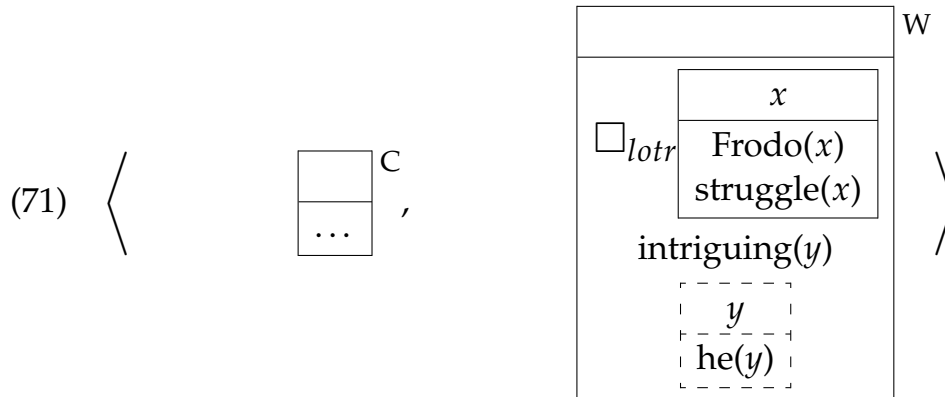
The pronoun 'he' in the metafictional statement in (67) triggers a presupposition that there is a masculine entity (denoted by the dashed box) and we update with the information that this masculine entity is an intriguing fictional character. The pronoun 'he' is anaphoric on the name 'Frodo' introduced in the preceding *parafictional* statement. Normally, we represent this by equating their discourse referents (i.e., resolving the presupposition and replacing all occurrences of y with x) so that the resulting update of the metafictional statement in (67) is 'intriguing(x)'. However, following standard DRT-rules, x is not accessible outside of the 'In *The Lord of the Rings*'-operator and hence the presupposition remains unresolved. The workspace account thus predicts that we cannot update the workspace with (and thus interpret) discourse involving metafictional anaphora such as in (67). Obviously, we *can* interpret (67) and hence we need to adjust our theory.

Related issues may arise with some cases of co-predication with mixed metafictional and parafictional discourse, i.e., when a fictional name is introduced in metafictional and parafictional discourse 'simultaneously'. In general terms, discourse referents introduced by parafictional or metafictional discourse have to be accessible to subsequent parafictional and metafictional statements to account for all admissible anaphoric dependencies. The DRT formalism thus shows that a dynamic approach that involves quarantining of fictional content under a fiction operator quickly runs into difficulties with discourse that involves anaphoric dependencies across metafictional discourse since metafictional statements are unembedded assertions. Although different solutions to this general issue are possible, we will see that the workspace account is pushed into the direction of a realist approach

where all discourse referents for fictional entities are moved to the main box to ensure they are always accessible.

6.3.2 Metafictional anaphora in the unofficial common ground accounts

The problem of accessibility introduced by metafictional anaphora generalizes to other current dynamic approaches to fiction that involve some type of embedding or separation of the content and discourse referents of the fictional narrative.[3] For instance, in Stokke's (2013; 2018) and in an Eckardt-style (2014) unofficial common ground account (see chapter 3), fictional and parafictional statements update a separated unofficial common ground related to the relevant fiction whereas metafictional statements, being a type of assertion, update the official common ground related to actual states of affairs. The complete common ground after updating with (67) will thus look as follows:

(72) $\left\langle \begin{array}{|l|} \hline \hspace{3cm} \text{Official} \\ \hline \text{intriguing}(y) \\ \begin{array}{|c|} \hline y \\ \hline \text{he}(y) \\ \hline \end{array} \\ \hline \end{array} \right. , \left. \begin{array}{|c|} \hline \hspace{2cm} \text{Unofficial}^{lotr} \\ \hline x \\ \hline \text{Frodo}(x) \\ \text{struggle}(x) \\ \hline \end{array} \right\rangle$

Although it is in principle formally possible to make the discourse referents of fictional entities accessible outside of the unofficial common grounds where they were introduced, this is inconsistent with the motivation for having separate unofficial common grounds for fiction. Namely, that the content of fictional narratives is somehow quarantined from information about actual states of affairs. Moreover, in Stokke's framework unofficial common grounds are – unlike the official common ground – temporary (i.e., they exist for the purpose and duration of the (para)fictional discourse). Making these temporary discourse referents accessible outside of the unofficial common grounds would lead to even more difficulties. Hence, in the unofficial common ground accounts, the presupposition triggered by the metafictional anaphora in (67) cannot straightforwardly be resolved because the discourse referent for 'Frodo' is not accessible outside of the unofficial common ground.

---

[3]With the possible exception of Maier's cognitive framework (see section 6.4.2).

6.4 A comparison of different solutions

In the rest of this chapter I describe and evaluate four different strategies to meet the described challenge of metafictional anaphora in the workspace account.[4] First, one can adopt a descriptivist approach and account for metafictional anaphora in non-dynamic terms (section 6.4.1). Alternatively, staying in the dynamic semantic framework, one can either stick to a descriptivist/anti-realist approach and adjust the accessibility relations (section 6.4.2) or accommodate a new discourse referent that is accessible through standard accessibility relations. Such a discourse referent can be understood as an abstract object in a realist framework (section 6.4.3) or as a dot-object (section 6.4.4). The proposed solutions are evaluated with respect to whether they can also be extended to account for parafictional anaphora or co-predication, and whether the solutions avoid the problem of the wrong kind of object in its different variants.

6.4.1 A descriptivist approach: A description of Frodo

D-type accounts

A possible solution to the described challenge in a traditional semantics framework is a descriptivist approach to anaphora (e.g., Evans (1977); Elbourne (2005); Heim (1990)). This analysis was originally proposed as a solution to the accessibility problem posed by donkey anaphora. Consider the following donkey sentence:

(73)   If Sarah owns a donkey, she beats it.

Intuitively, the pronoun 'it' does not refer to a particular individual donkey but is bounded by 'a donkey'. However, it is outside of the syntactic scope of 'a donkey' and hence inaccessible. On a descriptivist analysis, the anaphoric pronoun 'it' functions like, or 'goes proxy for', the definite description 'the donkey' retrieved from the preceding clause. In Elbourne's D-type account, this is because NPs at the level of syntax undergo phonetic deletion (are not pronounced at the surface level) when in the environment of an identical NP

---

[4]Prima facie, the discussed solutions can be extended to address the challenge in other dynamic approaches such as the unofficial common ground accounts. I leave exploration and evaluation of these parallel solutions to future research.

(e.g., 'My shirt is the same as his (shirt)'). Similarly, (73) is in fact equivalent to (74):

(74)   If Sarah owns a donkey, she beats the donkey.

This analysis evades the problem of the unbindable pronoun by replacing it with a definite description.

Addressing the challenge

When we apply this strategy to our central example of metafictional anaphora (67), the pronoun 'he' is also analysed as going proxy for a definite description retrieved from the previous clause. However, (67) cannot be the result of simple phonetic deletion of an identical NP. If it were, (67) would be equivalent to something like (75):

(75)   In *The Lord of the Rings*, Frodo goes through an immense mental struggle to save his friends. Ah yes, the person named Frodo in *The Lord of the Rings* that goes through an immense mental struggle to save his friends, is an intriguing fictional character!

This gives us an incorrect analysis of (67): A flesh and blood person cannot be a fictional character. In other words, through simple phonetic deletion we retrieve the 'wrong kind of definite description' from the parafictional statement and hence run into the anti-realist variant of the problem of the wrong kind of object.

To get the correct interpretation, what is required is a *metafictional* description such as 'the *character* named Frodo in *The Lord of the Rings*' so that (67) becomes equivalent to (76):

(76)   In *The Lord of the Rings*, Frodo goes through an immense mental struggle to save his friends. Ah yes, the fictional character named Frodo in *The Lord of the Rings* is an intriguing fictional character!

Although (76) gives an acceptable analysis of what is expressed by (67), it is unclear how to compositionally obtain such a meta-description of Frodo from the preceding clause. Moreover, even if we assume that we can accommodate such a definite description for metafictional anaphora, this solution does not extend to parafictional anaphora such as (68). In such cases both simple phonetic deletion and accommodation of a meta-description of Frodo lead to the same incorrect interpretation:

(77) Frodo is an intriguing fictional character. In *The Lord of the Rings*, the fictional character named Frodo goes through an immense mental struggle to save his friends.

Here we run into the realist variant of the problem of the wrong kind of object: a fictional character, being an abstract object, cannot go through an immense mental struggle to save his friends. Deriving the correct interpretation of (68) would thus require accommodation of yet another type of definite description, i.e., a *parafictional* description.

Last but not least, it is unclear how an account of phonetic deletion of NPs can be extended to cases of co-predication in mixed parafictional and metafictional discourse. For instance, in (69) there is no anaphoric pronoun that can be analysed as going proxy for a definite description.

Hence, a descriptivist approach does not (as yet) adequately solve the accessibility issues involved in mixed metafictional and parafictional discourse; simple phonetic deletion provides the wrong kind of definite descriptions in the case of metafictional anaphora and hence we need an account of how to accommodate the right kind of definite descriptions. This fix can, in turn, not be extended to parafictional anaphora or cases of co-predication.

6.4.2 Psychologistic DRT: Adjusting accessibility

Another possible solution to the problem of metafictional anaphora (that sticks to a dynamic approach of language and an anti-realist approach to fictional names) is to adjust the DRS accessibility relations so that the discourse referent for 'Frodo' *is* accessible.

Psychologistic DRT

Maier (2017) adopts this strategy in his psychologistic DRT framework in which the context that is updated by statements is an agent's mental state. The agent's mental state is represented as a set of DRS's that are linked to cognitive attitudes such as belief (BEL) and imagination (IMG). In line with the consensus view of fiction interpretation (e.g., Walton (1990); Currie (1990)), fictional discourse updates the imagination-box and non-fictional discourse updates the belief-box. Hence, assuming that the agent updates with (67) after having engaged with *The Lord of the Rings* – and thus having previously imagined Frodo – the agent's mental state after updating with (67), but before pronoun resolution, looks as follows:

(78)

$$\left\{ \left\langle \mathrm{IMG}, \begin{array}{|c|} \hline x \\ \hline \mathrm{Frodo}(x) \\ \ldots \\ \hline \end{array} \right\rangle, \left\langle \mathrm{BEL}, \begin{array}{|l|} \hline \Box_{lotr} \begin{array}{l} \mathrm{struggle}(y) \\ \begin{array}{|c|} \hline y \\ \hline \mathrm{Frodo}(y) \\ \hline \end{array} \\ \mathrm{intriguing}(z) \\ \begin{array}{|c|} \hline z \\ \hline \mathrm{he}(z) \\ \hline \end{array} \end{array} \\ \hline \end{array} \right\rangle \right\}$$
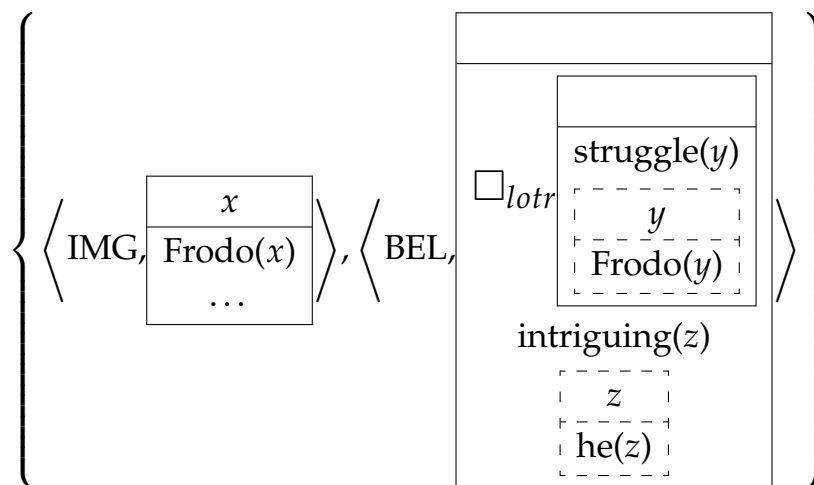
The parafictional and metafictional statement function similarly in that both, being assertions, update the belief-box and trigger presuppositions that need to take as their discourse referent x for 'Frodo'.[5]
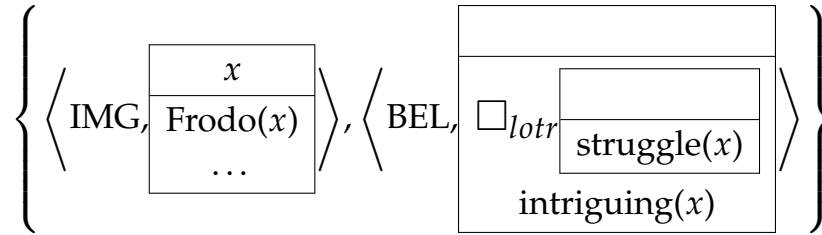
Although this dynamic semantic framework – like the workspace account and the unofficial common ground accounts – involves separating the discourse referents of fictional entities, there is no accessibility problem because Maier assumes, contra usual practice, that attitudes can be referentially dependent on attitudes other than belief.[6] He gives the example of someone who wants to buy a new smartphone in a few years and imagines it to have a flexible transparent screen. This is a desire dependent imagination. Similarly, doxastic attitudes can be referentially dependent on imagination. When engaging in *The Lord of the Rings* I imagine the existence of an entity named Frodo and when engaging in parafictional or metafictional discourse such as (67) I believe that in *The Lord of the Rings* this entity went through a mental struggle and that this entity is an intriguing fictional character.[7] Hence the presuppositions triggered by the metafictional anaphora in (67) can be resolved; the content of the imagination-box is accessible:

---

[5]Here I follow Geurts (1997) in analysing proper names as triggering presuppositions.

[6]Maier does allow that there may turn out to be some structural constraints on specific cross-attitudinal dependencies.
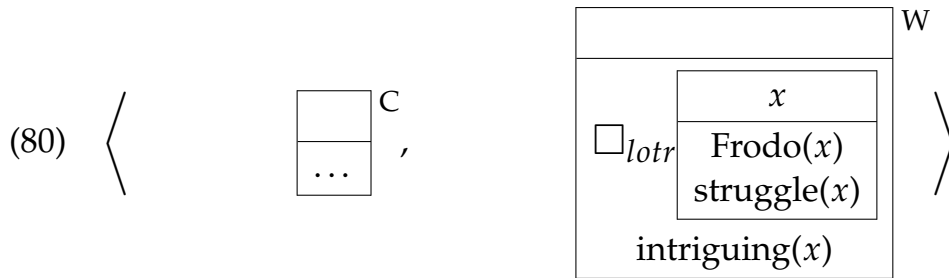
[7]Maier thus offers a uniform anti-realist semantic treatment of fictional names across parafictional, metafictional *and* fictional discourse. All different types of discourse operate on the same discourse referents.

(79)

$$\left\{ \left\langle \text{IMG}, \boxed{\begin{array}{c} x \\ \hline \text{Frodo}(x) \\ \dots \end{array}} \right\rangle, \left\langle \text{BEL}, \boxed{\begin{array}{c} \Box_{lotr} \boxed{\text{struggle}(x)} \\ \hline \text{intriguing}(x) \end{array}} \right\rangle \right\}$$

Addressing the challenge

Extending this strategy to the workspace account would amount to changing the accessibility relations relative to the content embedded under the 'In fiction s'-operator so that the presupposition triggered by the pronoun 'he' in the metafictional statement in (67) can be resolved:

(80)

$$\left\langle \boxed{\begin{array}{c} \\ \hline \dots \end{array}}^{C}, \boxed{\begin{array}{c} \hfill W \\ \hline \Box_{lotr} \boxed{\begin{array}{c} x \\ \hline \text{Frodo}(x) \\ \text{struggle}(x) \end{array}} \\ \hline \text{intriguing}(x) \end{array}} \right\rangle$$

We can thus stick to the basic anti-realist version of the workspace account: The name 'Frodo' refers in both parafictional and metafictional discourse to a flesh and blood entity that only exists in *The Lord of the Rings* worlds. Such a move would solve the accessibility issues involved in metafictional anaphora. The solution can also be extended to account for parafictional anaphora and co-predication in mixed parafictional and metafictional discourse. Discourse referents for fictional entities are always embedded under fiction operators in the common ground. But since such discourse referents are accessible outside of the embedded DRS, presuppositions that are triggered by subsequent or simultaneous metafictional discourse can always be resolved.

However, although there may be independent reasons to make the content in the imagination-box accessible in Maier's cognitive framework, there are none for doing this with the 'In fiction s'-operator, making this an *ad hoc* move. More importantly, there are already substantial theoretical costs in Maier's cognitive framework to making the content of the imagination-box accessible (i.e., a highly complex semantic system) but such a move with

the 'In fiction s'-operator would amount to a drastic change of the basic semantics of DRT (in which any type of embedding entails inaccessibility). Such a radical departure of standard DRT semantics is undesirable.

In fact, the accessibility problem of metafictional anaphora may reappear in the psychologistic DRT framework as well. Because Maier assumes that parafictional and metafictional statements are referentially dependent on existential imagination (induced by fictional statements), he only considers discourse in which the interpretation of parafictional and metafictional statements comes *after* the interpretation of fictional statements. However, fictional names can also be introduced in parafictional or metafictional statements. Suppose (67) featured the first occurrence of the name 'Frodo'. In this case the fictional name 'Frodo' introduced in the parafictional statement cannot be referentially dependent on a previous act of imagination. In Maier's framework, interpretation of such discourse would have to involve either accommodation of a kind of contentless or minimal imagination during the parafictional discourse (e.g., imagining that there is a person named Frodo) or involve local accommodation of a discourse referent in the belief-box. The latter strategy would result in the following representation of the agent's mental state:

(81)
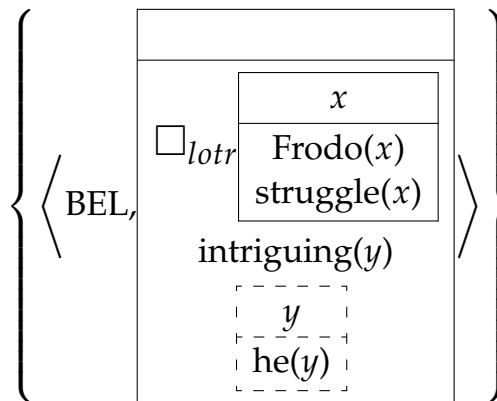
Here the discourse referent for 'Frodo' is embedded in the 'In *The Lord of the Rings*'-operator rather than the imagination-box. Making this discourse referent accessible will lead to the aforementioned problems of changing the accessibility relations relative to the 'In fiction s'-operator.

In any case, a solution based on changing the accessibility relations plainly runs into the anti-realist variant of the problem of the wrong kind of object. The imagined entity (in terms of the psychologistic DRT framework) or
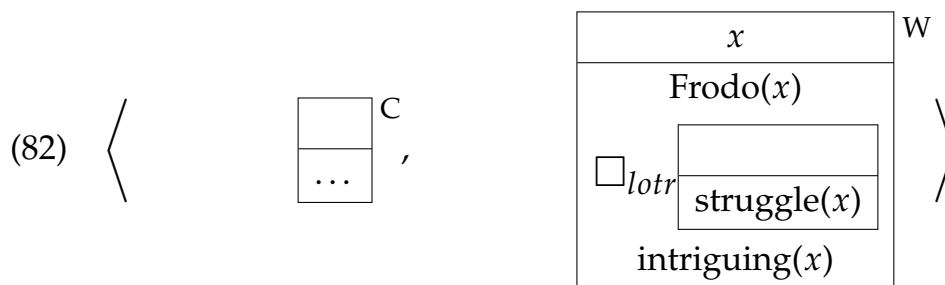
the entity that exists in the *The Lord of the Rings* worlds (in terms of the workspace account) is a flesh and blood hobbit. If we make the discourse referent for this object accessible to the metafictional statement in (67), the metafictional statement will be about this flesh and blood individual. However, a flesh and blood hobbit cannot be an intriguing fictional character.

Hence, a solution based on Maier's psychologistic DRT does not adequately solve the problem of metafictional anaphora. It cannot straightforwardly be extended to a workspace account and runs into the anti-realist version of the problem of the wrong kind of object.

6.4.3 Object theory: Frodo the abstract object

An alternative strategy to address the challenge of metafictional anaphora is to claim that fictional names in parafictional and metafictional statements refer to an object with a discourse referent that is accessible in the main box:

$$(82) \quad \left\langle \quad \boxed{\begin{array}{c} \\ \hline \dots \end{array}}\text{C} \quad , \quad \boxed{\begin{array}{c} x \quad \text{W} \\ \hline \text{Frodo}(x) \\ \Box_{lotr}\boxed{\text{struggle}(x)} \\ \text{intriguing}(x) \end{array}} \quad \right\rangle$$

The above DRS formally resembles a DRS you would expect on a de re version of the workspace account (see section 4.6.1) where fictions that are about non-fictional entities (e.g., Napoleon in *War and Peace*) are analysed as being de re about these entities (rather than about their fictional substitutes). Parafictional discourse about such fictions (e.g., 'In *War and Peace*, Napoleon examines the Pratzen Heights') thus takes the discourse referent of an actually existing individual from the main box. However, contrary to Napoleon, Frodo never really existed. One way to account for this is to follow Zalta's application of his logic of abstract objects – 'object theory' (Zalta (1983, 1988)) – to fiction and claim that parafictional and metafictional statements are about abstract objects (e.g., Frodo the fictional character) that really exist.[8]

---

[8]For details on the object theoretic treatment of fiction see especially Zalta's (1983) chapter IV, (1988) chapter 7 and Zalta (2000; 1987]).

6 The challenge of metafictional anaphora

Object theory

Prima facie, this strategy seems to run head first into the realist version of the problem of the wrong kind of object: both the parafictional and metafictional statements in (67) operate on a discourse referent for an abstract object, but surely an abstract object cannot go through an immense mental struggle! To avoid this problem Zalta distinguishes between two types of objects and two modes of predication.[9] x is an 'ordinary object' ('O!(x)') if it is, or could have been, concrete (e.g., Mount Everest). x is an 'abstract object' ('A!(x)') just in case it isn't and could not have been concrete (e.g., the empty set). An ordinary object like Mount Everest can 'exemplify' being 8,848 meters tall, i.e., it has the property of being 8,848 meters tall in the standard sense. Zalta denotes this as it would be denoted in standard predicate logic, i.e., with the argument to the right of the predicate. Hence, a natural language statement such as (83) is translated into the formal language of object theory as follows:

(83) Mount Everest is 8,848 meters tall.

(84) 8,848m-tall(m)[10]

In contrast, an abstract object can 'encode' a property which means it has this property as one of its constitutive characteristics. What properties an abstract object encodes are the properties that *define* the abstract object. For instance, the empty set encodes the property of having no members. This is denoted with the argument to the left of the predicate. Hence, a natural language statement such as (85) is translated as follows:

(85) The empty set has no members.

(86) $(\varnothing)$memberless

Ordinary objects cannot encode properties but abstract objects *can* exemplify properties. For instance, the empty set exemplifies (but does not encode) being widely discussed; it has this property but it is not a property that is constitutive of its essence. A statement such as (87) is thus translated as follows:

(87) The empty set is widely discussed.

---

[9]This distinction originally comes from Mally (1912). See also van Inwagen (2000).

[10]In fact, Zalta uses standard predicate logic notation (i.e., (83) is translated as 'Tm'). The current notation is compatible with the standard notation for DRS conditions.

(88) widely-disc($\varnothing$)

Two types of abstract objects are relevant for the object theoretic analysis of fictional, parafictional and metafictional statements: 'stories' and 'fictional characters'. A story (e.g., *The Lord of the Rings*) is an abstract object that encodes the content of a narrative. It encodes 'vacuous' or 'propositional' properties of the form 'being such that ϕ is true', where ϕ is a proposition that is true in the story. A fictional character is an abstract object that is *native* to a story (e.g., Frodo or the One Ring, but not Napoleon).

The abstract object theoretic analysis of parafictional discourse differs substantially from the Lewisian analysis. Contrary to common practice, Zalta draws a strong distinction between the analysis of explicit parafictional statements such as (58) and implicit parafictional statements such as (89):

(89) Frodo was born in the Shire.

This is because Zalta is a realist about fictional characters (i.e., they exist as abstract objects) and hence we can talk about them as we do about ordinary objects (i.e., without an 'In fiction *s*'-operator or some type of pretence). A statement such as (89) is thus actually not 'implicit' in the sense that it is covertly embedded. Rather, it is a plain statement about what properties a certain abstract object encodes:

(90) (f)born.in.Shire

Explicit parafictional statements (e.g., (58)) on the other hand do contain an 'In fiction *s*'-operator (i.e., Σ_s). However, this is not a classical Lewisian fiction operator (i.e., $\square_s$). Explicit parafictional statements are statements about specific encoding and exemplifying relations between stories and characters. For instance, (58) expresses that *The Lord of the Rings encodes* the vacuous property of being such that Frodo exemplifies living in the Shire:

(91) Σ_{lotr}born.in.Shire(f)

Zalta proves a theorem in his theory:

(92) $\forall x \forall s(\text{Native}(x, s) \rightarrow \forall F(xF \equiv \Sigma_s Fx))$

according to which, if some character x is native to some story s, implicit and explicit parafictional statements about x (in s) (e.g., (58) and (89)) are equivalent.

Metafictional statements are statements about what properties fictional characters *exemplify*. For instance, the metafictional statement in (67) ex-

presses that Frodo exemplifies the property of being an intriguing fictional character:

(93)   intriguing(f)

In line with the desideratum formulated in section 6.2, object theory does not extend the uniform semantic treatment of fictional names across fictional discourse. Fictional statements do not involve reference to abstract objects but rather constitute the practice of story telling that determines – through an extended 'naming baptism' – what abstract objects the fictional names in parafictional and metafictional statements refer to (see Zalta (2000, 1987); Semeijn and Zalta (2021)). For instance, through Tolkien's act of writing *The Lord of the Rings*, the unique abstract object that the name 'Frodo' refers to (in parafictional and metafictional discourse) is determined. Similar accounts of reference to fictional characters supervening on fictional discourse can be found in Kripke (1973), Schiffer (2003) and Searle (1975) (but see also Hunter (1981) for a contrary view).
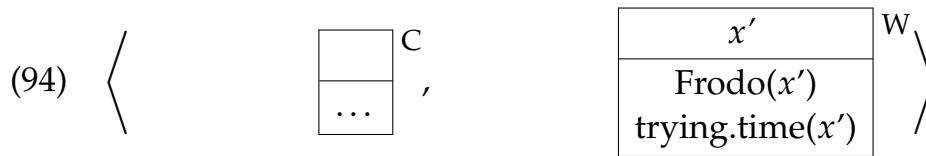
Fictive closure*

If we incorporate Zalta's analysis of fictional names into the workspace account, we need to replace our Lewisian analysis of parafictional updates of the common ground by the object theoretical analysis. This entails a different analysis of parafictional discourse but also entails a modification of the fictive closure operation (which also involves a parafictional update). First and foremost, we replace the Lewisian operator ($\Box_s$) with the object theoretic fiction operator (Σ_y). Second, because of the strong distinction drawn between implicit and explicit parafictional statements, fictive closure can in theory involve two different kinds of updates of the common ground: an update with explicit parafictional information and, if the relevant fictional characters are native to the relevant story, also with implicit parafictional information.[11] The following is a representation of fictive closure* of (25) that includes updates of the common ground with both types of statements. First, a workspace is updated with content of the fictional statement, i.e., it is temporarily commonly accepted that there is an entity called Frodo that had a very trying time some afternoon:
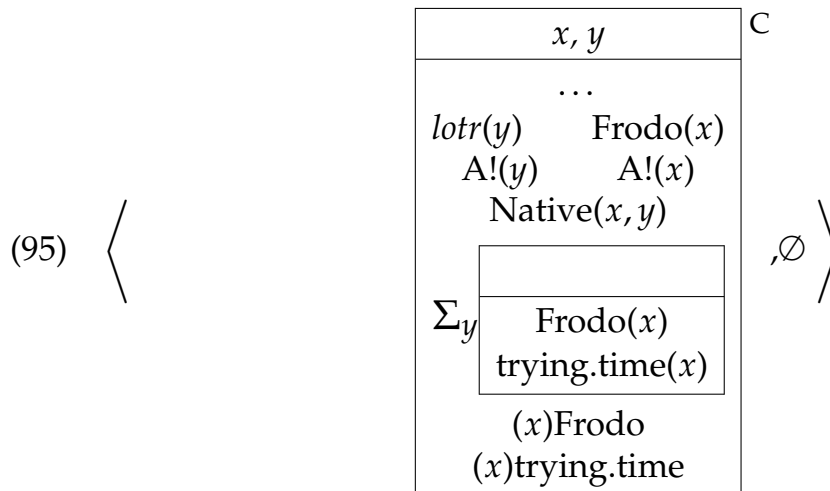
---

[11]If we also incorporate theorem (92) we can simplify the representation of the common ground with respect to statements about native fictional characters.

(94) $\left\langle \quad \boxed{\begin{array}{c} \\ \hline \cdots \end{array}}\,C \;,\; \boxed{\begin{array}{c} x' \\ \hline \text{Frodo}(x') \\ \text{trying.time}(x') \end{array}}\,W \quad \right\rangle$

At fictive closure*, as soon as we stop engaging in the fictional discourse, we update the common ground with discourse referents for the newly introduced abstract objects (e.g., the story *The Lord of the Rings* and the fictional character Frodo) and with (explicit and implicit) parafictional information based on the content of the workspace (e.g., 'Σ_{lotr}trying.time(x)' and '(x)trying.time'):

(95) $\left\langle \quad \boxed{\begin{array}{c} x,\,y \qquad\quad C \\ \hline \cdots \\ lotr(y) \qquad \text{Frodo}(x) \\ A!(y) \qquad A!(x) \\ \text{Native}(x,y) \\ \Sigma_y\;\boxed{\begin{array}{c} \\ \hline \text{Frodo}(x) \\ \text{trying.time}(x) \end{array}} \\ (x)\text{Frodo} \\ (x)\text{trying.time} \end{array}} \;,\varnothing \quad \right\rangle$
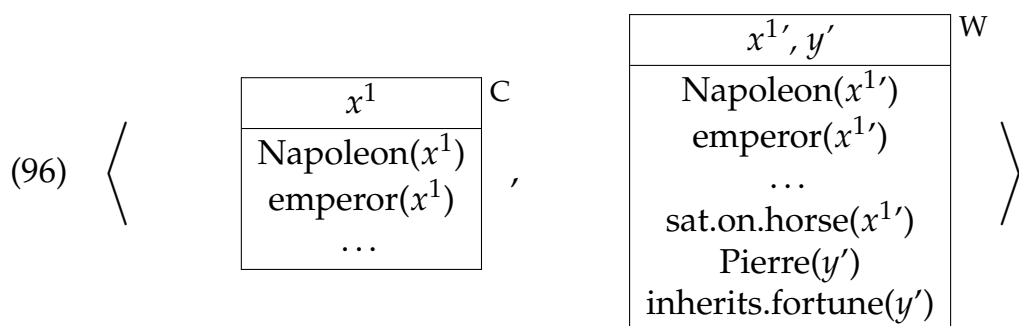
Importantly, not all propositional content of the workspace is updated as parafictional information simpliciter: proper name conditions (e.g., 'Frodo(x)') are doubled and placed in the main box. This represents the fact that the abstract object Frodo also *exemplifies* being named 'Frodo' outside of *The Lord of the Rings*.

This move comes at a significant theoretical cost since it greatly complicates fictive closure (and opening). The name-predicate 'Frodo' now occurs three times in the DRS, which is counterintuitive. More importantly, some discourse referents for ordinary objects in the workspace (e.g., Frodo, Moria, the One Ring) are replaced with discourse referents for abstract objects in the main box (and vice versa for fictive opening). However, discourse referents for ordinary things in the workspace that have non-fictional substitutes in the common ground (e.g., Napoleon, Paris, the British crown jewels) stay embedded under the fiction operator. The analysis is, however, in line with Zalta's analysis of fictional discourse (i.e., updating a workspace and performing fictive closure) as an extended naming baptism that reference to
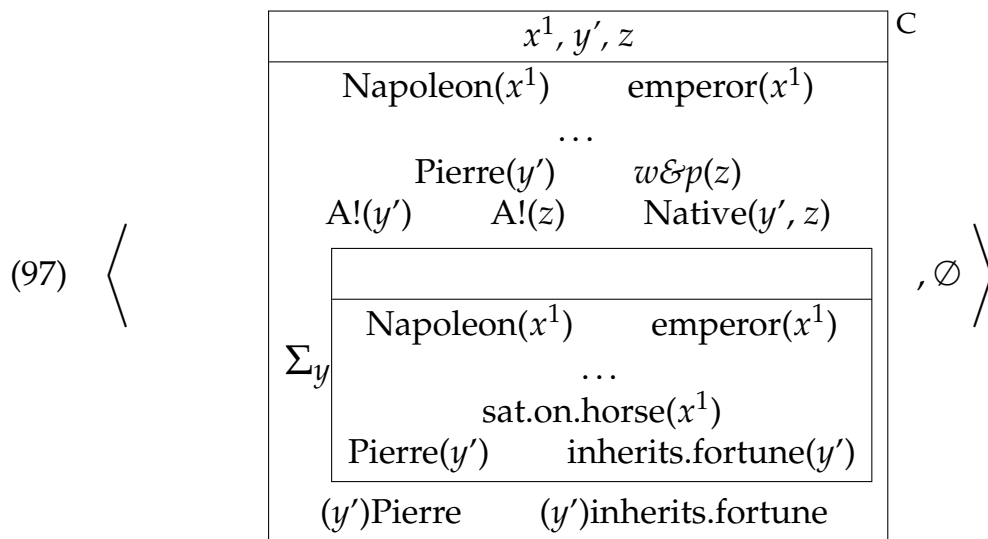
really existing fictional characters (i.e., entities that are represented in the main DRS) supervenes on.

In fact, we can design a more uniform fictive closure* operation if we combine the object theoretic approach to fictional names with a de re version of the workspace account (see section 4.6.1) where fiction about non-fictional entities (e.g., Napoleon in *War and Peace*) is de re about these entities. On such an account, all discourse referents for non-fictional entities in the common ground are anchored to their copies in the workspace. These copies 'move out' of the DRS that is embedded under a fiction operator at fictive closure and are replaced by the 'original' discourse referents for these non-fictional entities. Combining the object theoretic account of fictional names and a de re account of fiction about non-fictional entities results in a uniform fictive closure operation where *all* discourse referents (whether for non-fictional entities or fictional entities) 'move out' of the embedded DRS. Those that are anchored to already existing discourse referents in the main DRS (i.e., discourse referents for non-fictional entities or discourse referents for previously introduced fictional characters) are replaced by those, those that are not (i.e., discourse referents newly introduced in the workspace) give rise to new abstract objects. Consider for instance fictive closure after reading about Napoleon and the fictional character Pierre in *War and Peace*. Suppose that it was already common ground that Napoleon was a French emperor and that the workspace has been updated with (amongst other things) the information that Napoleon sat on an small gray Arab horse in front of his Marshals and that there was a guy named Pierre who inherited a great fortune:

$$
(96) \quad \left\langle \quad
\boxed{\begin{array}{c} x^1 \quad\quad \text{C} \\ \hline \text{Napoleon}(x^1) \\ \text{emperor}(x^1) \\ \ldots \end{array}} \quad , \quad
\boxed{\begin{array}{c} x^{1\prime}, y' \quad\quad \text{W} \\ \hline \text{Napoleon}(x^{1\prime}) \\ \text{emperor}(x^{1\prime}) \\ \ldots \\ \text{sat.on.horse}(x^{1\prime}) \\ \text{Pierre}(y') \\ \text{inherits.fortune}(y') \end{array}} \quad \right\rangle
$$

At fictive closure all discourse referents now move out of the embedded DRS:

(97)
$$\left\langle \; \begin{array}{|c|}\hline x^1,\, y',\, z \\\hline \mathrm{Napoleon}(x^1) \qquad \mathrm{emperor}(x^1) \\ \cdots \\ \mathrm{Pierre}(y') \qquad w\&p(z) \\ \mathrm{A!}(y') \qquad \mathrm{A!}(z) \qquad \mathrm{Native}(y',z) \\ \Sigma_y \begin{array}{|c|}\hline \\\hline \mathrm{Napoleon}(x^1) \qquad \mathrm{emperor}(x^1) \\ \cdots \\ \mathrm{sat.on.horse}(x^1) \\ \mathrm{Pierre}(y') \qquad \mathrm{inherits.fortune}(y') \\\hline \end{array} \\ (y')\mathrm{Pierre} \qquad (y')\mathrm{inherits.fortune} \\\hline \end{array}^{\;C} \;,\, \varnothing \; \right\rangle$$

I leave further exploration into the merits of combining fictive closure* with a de re analysis to future research.[12] The rest of this section deals with the object theoretic solution to metafictional anaphora which is compatible with a de re version of the workspace account and a version that is (apart from its treatment of fictional names) descriptivist.

Addressing the challenge

If we adopt the object theoretic strategy we add an abstract object to the shared ontology for any fictional entity that is introduced and is native to the relevant story. This means that we incorporate Zalta's metaphysical assumptions that entail the existence of abstract objects in the actual world. It also means that after a parafictional update (e.g., after engaging in parafictional or fictional discourse) the discourse referent for (the abstract object) 'Frodo'

---

[12] Arguably, the described account also offers a natural analysis of anaphoric dependencies across parafictional statements and 'regular' assertions about non-fictional entities. Consider:
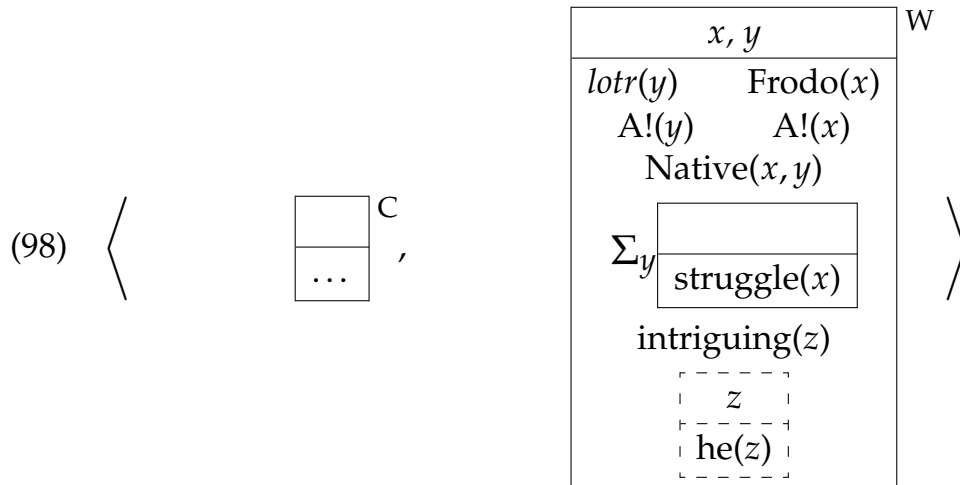
   (iii)   In *War and Peace*, Napoleon$_i$ is a hero. But actually, he$_i$ was an overconfident opportunist.

Anaphoric dependencies as in (iii) may lead to comparable accessibility issues as metafictional anaphora on a simple descriptivist version of the workspace account (in this case there *is* in fact a discourse referent for Napoleon accessible in the main DRS but it is not the same as the one that is referred to in the parafictional statement). However, for these cases there is no additional problem of the wrong kind of object looming in the background. I leave exploration of these issues to further research.

is accessible outside of the 'In *The Lord of the Rings*'-operator. To see how this solves the challenge posed by metafictional anaphora, we first have to recognize that because Zalta draws a strong distinction between implicit and explicit parafictional statements, the challenge splits up in two sub-challenges: One of pronominal anaphora across mixed *explicit* parafictional and metafictional discourse and one of pronominal anaphora across mixed *implicit* parafictional and metafictional discourse. The central example up to this point, (67), is an example of the first kind. I represent the workspace updated with (67) as follows:

(67)   In *The Lord of the Rings*, Frodo$_i$ goes through an immense mental struggle to save his$_i$ friends. Ah yes, he$_i$ is an intriguing fictional character!

(98)
$$\left\langle \quad \boxed{\begin{array}{c} \\ \hline \dots \end{array}}\,C \quad , \quad \boxed{\begin{array}{c} x, y \\ \hline lotr(y) \qquad \text{Frodo}(x) \\ \text{A!}(y) \qquad \text{A!}(x) \\ \text{Native}(x, y) \\ \Sigma_y \boxed{\begin{array}{c} \\ \hline \text{struggle}(x) \end{array}} \\ \text{intriguing}(z) \\ \boxed{\begin{array}{c} z \\ \hline he(z) \end{array}} \end{array}}\,W \quad \right\rangle$$
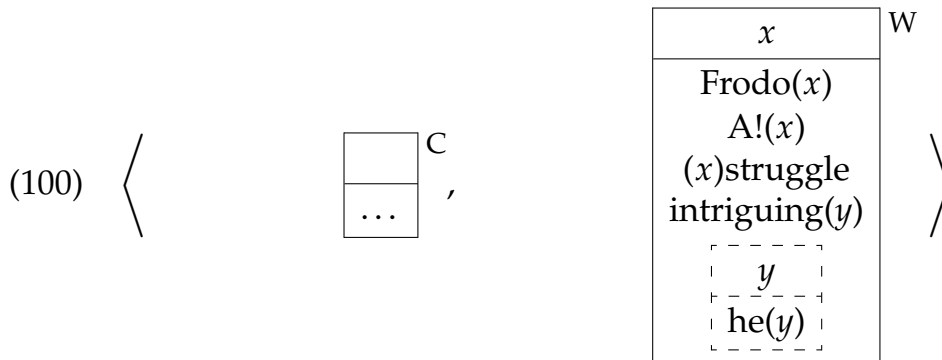
Next, we can make a variation of (67) so that it is an example of pronominal anaphora across mixed *implicit* parafictional/metafictional discourse:

(99)   Frodo$_i$ goes through an immense mental struggle to save his$_i$ friends. Ah yes, he$_i$ is an intriguing fictional character!

I represent the workspace updated with (99) as follows:

(100) $\left\langle \quad \boxed{\begin{array}{c} \\ \hline \dots \end{array}}\,C \;,\; \boxed{\begin{array}{|c|}\hline x \quad\quad W \\ \hline \text{Frodo}(x) \\ \text{A!}(x) \\ (x)\text{struggle} \\ \text{intriguing}(y) \\ \hline \begin{array}{|c|}\hline y \\ \hline he(y) \\ \hline\end{array} \\ \hline \end{array}} \quad \right\rangle$

As the formulas show, in both cases the discourse referent *x* for 'Frodo' is accessible outside of the 'In *The Lord of the Rings*'-operator. Hence the presupposition triggered by the pronoun 'he' in the metafictional statement in (67) and (99) (an assertion about what properties the abstract object Frodo *exemplifies*) *can* be resolved.[13]

This approach easily extends to parafictional anaphora and cases of co-predication in mixed metafictional and parafictional discourse. In general, metafictional, explicit parafictional and implicit parafictional discourse all operates on the same globally accessible discourse referents for fictional characters and hence any kind of mixed discourse will be interpretable.

Revisiting the problem of the wrong kind of object

Although this analysis seems to straightforwardly solve the problem of metafictional anaphora while avoiding the problem of the wrong kind of object for *implicit* parafictional statements (i.e., abstract objects are the 'right kind of objects' to *encode* properties such as living in the Shire) and metafictional statements (i.e., abstract objects are the 'right kind of objects' to *exemplify* properties such as being an intriguing fictional character), the problem seems to reappear for the object theoretic treatment of *explicit* parafictional statements. Recently, Klauk (2014) has argued along these lines. The worry is that because fictional names like 'Frodo' refer uniformly to abstract objects in metafictional and parafictional discourse, we would be

---

[13]In (98) and (100) I assume the standard analysis of the pronoun 'he' as triggering a presupposition that there is a *masculine* entity. However, although abstract objects such as Frodo may *encode* masculinity, they cannot *exemplify* this property; abstract objects have no gender. To make the object theoretic solution work we have to assume that the pronoun 'he' triggers the presupposition of the existence of an object that is masculine in an underspecified way.

able to infer something like (101) from an explicit parafictional statement such as (58):

(58)   In *The Lord of the Rings*, Frodo was born in the Shire.

(101)   In *The Lord of the Rings*, an abstract object was born in the Shire.

(101) is obviously problematic because it leads to a reoccurrence of the realist variant of the problem of the wrong kind of object: *The Lord of the Rings* is a story about flesh and blood hobbits, not about what properties abstract objects exemplify. Moreover, although we may be able to imagine that an abstract object *exemplifies* living in the Shire, this amounts to imagining a category mistake (i.e., abstracts object cannot live in certain regions) which should be unusual and remarkable and cannot comprise our common practice of engaging in the content of a fictional work.

Semeijn and Zalta (2021) offer a formalisation of Klauk's objection and divide it into three sub-challenges which they address separately. Below I briefly discuss the first and third sub-challenge they discuss.

First, the fact that f in (91) (and hence the name 'Frodo' in the explicit parafictional statement (58)) refers to an abstract object may raise the following worry: (91) expresses that *The Lord of the Rings* (an abstract object) encodes the vacuous property of being such that Frodo (another abstract object) exemplifies living in the Shire. In object theory, what's encoded by a fictional story is supposedly what is true in the fiction. Moreover, what's true in a fiction is what we are mandated to imagine when we engage with the fiction. But then object theory would imply that we are mandated to imagine something about an abstract object! In fact, this reasoning is flawed. In the object theoretic version of the workspace account, reference to abstract objects in explicit parafictional statements does *not* imply that we are mandated to imagine anything about abstract objects. Insofar as a fiction's 'mandate to imagine' plays a role in the workspace account (or in object theory), it is something that is triggered by fictional discourse and not by parafictional discourse. What we are 'mandated to imagine when engaging in a fiction' is what we temporarily accept or entertain while updating the workspace with the relevant fictional discourse. In the case of *The Lord of the Rings*, the workspace is updated with information about flesh and blood hobbits and their adventures. As fictive closure* shows, once we stop engaging with the fiction (and hence stop imagining), this previously imagined information gets converted into parafictional updates of the common ground that make reference to abstract objects. Parafictional

discourse, being non-fictional statements, operate on this information directly through assertive closure. In other words, even though parafictional statements do track or 'echo' the fictional discourse (and hence are reports on what is true in fiction), parafictional discourse itself is non-fictional discourse about abstract objects. This does not mean, however, that updates of the workspace triggered by fictional discourse also make reference to abstract objects. When engaging in *The Lord of the Rings*, we temporarily imagine things about Frodo the flesh and blood hobbit. Afterwards, we permanently believe things about the abstract object Frodo.

Even if reference to abstract objects in explicit parafictional statements such as (58) is not problematic, a further worry is whether – given that *f* in (91) refers to an abstract object – we can derive problematic statements (e.g., something like (101)) in object theory, as Klauk seems to think. One way this worry may be developed is by looking at the identification conditions of the abstract object called 'Frodo'. Given that Frodo is native to *The Lord of the Rings*, and given theorem (92), we can derive that Frodo is the abstract object that encodes all properties that Frodo exemplifies in *The Lord of the Rings*:

(102) $\quad f = \imath x(A!(x) \wedge \forall F((x)F \equiv \Sigma_{lotr}F(f)))$

Given that the formal language of object theory is completely extensional or 'denotational' (i.e., substitution of co-referring expressions is licensed in any context), we are licensed to substitute the definite description in (102) for *f* in (91). We can thus derive:

(103) $\quad \Sigma_{lotr}\text{born.in.Shire}(\imath x(A!(x) \wedge \forall F((x)F \equiv \Sigma_{lotr}F(f))))$

Reading (103) back into natural language would give us the following parafictional discourse:

(104) In *The Lord of the Rings*, the abstract object that encodes exactly those properties that Frodo exemplifies in *The Lord of the Rings*, was born in the Shire.

Surely, our theory should not predict that an utterance of (104) is licensed in a parafictional discussion on *The Lord of the Rings* because (104) ascribes inappropriate content to *The Lord of the Rings*. In fact, object theory does not predict this. Although (103) *is* a theoretical consequence of the theory, it cannot be read back into the language of the parafictional data. (104) includes technical terms such as 'encoding' and 'exemplifying' which were

introduced in object theory to disambiguate the two modes of predication that natural language conflates. In other words, although (104) may be correctly used by an object theorist engaged in a theoretical discussion on the abstract objects *The Lord of the Rings* and Frodo, the technical machinery of object theory used in (104) cannot simply be read back into natural language, i.e., the data we are trying to explain.[14]

Lastly, we can develop the worry one step further as follows: even if (103) can't be read back into natural language as (104) – unless as a part of an object theoretic discussion – we can potentially derive something even more problematic from it. Object theory includes the principle *Descriptions* (1988, p. 90), which is a version of Russell's (1905) Theory of Descriptions:

(105)   (a) $P\imath x(Qx) \equiv \exists! y Qy \wedge \exists y(Qy \wedge Py)$
        (b) $\imath x(Qx)P \equiv \exists! y Qy \wedge \exists y(Qy \wedge yP)$

In words, a formula of the form 'the Q exemplifies/encodes property P' is equivalent to 'there is a unique y that exemplifies Q and there is a y that exemplifies Q and that also exemplifies/encodes P. If we could apply *Descriptions* to the formula embedded under the Σ_{lotr} operator in (103), we would be able to derive the following:

(106) $\Sigma_{lotr} \exists! y(A!y \wedge \forall F(yF \equiv \Sigma_{lotr}Ff)) \wedge$
        $\exists y(A!y \wedge \forall F(yF \equiv \Sigma_{lotr}Ff) \wedge Sy)$

Formula (106) says (amongst other things) that *The Lord of the Rings* encodes that there is an abstract object that was born in the Shire. Surely, even someone engaged in an object theoretic discussion should not be licensed to say that. The abstract object *The Lord of the Rings* does not encode any existence claims about abstract objects. In fact, this inference is *not* licensed because, although the encoding environment created by the fiction operator may be denotational, it is also *hyperintensional*, i.e., it does not allow for substitution

---

[14]The second sub-challenge discussed by Semeijn and Zalta (2021) relates to substitutions of definite descriptions in the formal language that *can* be read back into natural language. Consider for instance:

   (iv)   In *The Lord of the Rings*, the character portrayed by Elijah Wood was born in the Shire.

Object theory predicts that (iv) is part of the parafictional data in natural language. Semeijn and Zalta suggest that this is not problematic because explicit parafictional statements are subject to a de re/ de dicto ambiguity and (iv) is in fact true on a de re reading.

of necessary equivalents salva veritate. This is possible because, in object theory, necessary equivalence does not entail identity (cf. Myhill (1963)) (e.g., the property of 'being a brown and colourless dog' is necessary equivalent to the property of 'being a barber that shaves all and only those that don't shave themselves' but these are distinct properties). Hence the inference from our application of *Descriptions* to the formula embedded under Σ_{lotr} in (103), to (106) is not licensed; we cannot substitute the equivalent formulas within the hyperintensional environment.

Outlook

Although we can incorporate the object theoretic analysis of metafictional and parafictional discourse into the workspace account (and adjust our fictive closure operation accordingly) and hence arrive at a consistent account of the different anaphoric dependencies that are possible across mixed discourse (that avoids the problem of the wrong kind of object), it does come at some theoretical costs. The main issue in combining object theory with the workspace account is that we are forced to give up the Lewisian analysis of the fiction operator in favour of the object theoretic analysis of the fiction operator. First, on the earlier adopted Lewisian analysis of parafictional discourse, implicit parafictional discourse is covertly prefixed by a fiction operator. This is inconsistent with the object theoretic analysis of implicit parafictional statements as unprefixed statements about what properties fictional characters *encode*. Second, even if we wanted to maintain the Lewisian analysis only for explicit parafictional statements, reference to abstract objects in explicit parafictional statements then does become problematic. Assuming that 'being abstract' is an essential property in Kripke's (1980) sense (i.e., a property that cannot change across possible worlds) and given that Frodo is an abstract object in the real world, Frodo will be an abstract object in all possible worlds including the *The Lord of the Rings* worlds. Arguably, we would thus be able to derive that in *The Lord of the Rings*, an abstract object was born in the Shire.

Having to give up the Lewisian analysis of parafictional discourse is somewhat disappointing. First, we lose the benefit of the very natural tie between the common ground framework and the Lewisian analysis that defines fictional truth in terms of overt beliefs in the community of origin. Second, the object theoretic analysis of parafictional discourse applies to both statements of the form 'In *s*, *ϕ*' and 'According to *s*, *ϕ*'. We thus lose

any explanatory value of the two offered semantic analyses of 'In' and 'According to' concerning the diverging linguistic behaviour of these operators (see chapter 7).
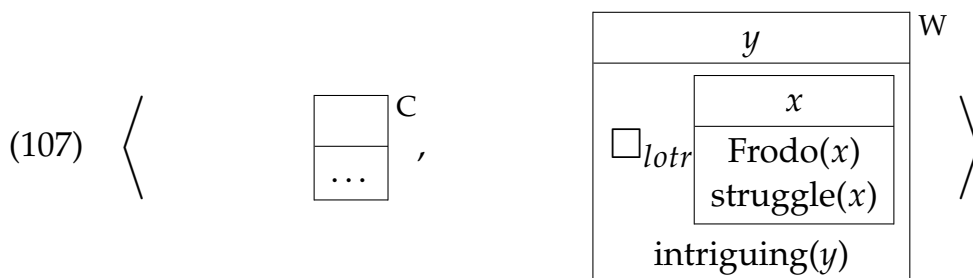
Hence, although an abstract object account adequately addresses the challenge of metafictional anaphora while avoiding the problem of the wrong kind of object, the account only works properly if we give up the Lewisian analysis of parafictional discourse. If we want to hold on to the Lewisian analysis, fictional names in parafictional discourse will somehow need to refer to flesh and blood individuals rather than abstract objects. Hence, in order to account for anaphoric dependencies across mixed metafictional and parafictional discourse while still avoiding the problem of the wrong kind of object in its different variants, it seems we will need to posit some kind of ambiguity in fictional names. The next section explores such an account.

6.4.4 Dot-object theory: The different facets of Frodo

An alternative solution to the problem of metafictional anaphora that is formally similar to the object theoretic solution but that incorporates an ambiguity analysis of fictional names, is to follow Recanati (2018) and claim that fictional names in parafictional and metafictional statements refer to so-called 'dot-objects' that are accessible in the main box.

Dot-object theory

Recanati goes back to Kripke's and Currie's intuition that fictional names are ambiguous. In parafictional statements 'Frodo' refers to a flesh and blood individual and in metafictional statements 'Frodo' refers to an abstract object. Such an approach to our central example of metafictional anaphora (67) would result in the following workspace update:

$$(107) \quad \left\langle \quad \boxed{\begin{array}{c} \\ \hline \dots \end{array}}^{\text{C}}, \quad \boxed{\begin{array}{c} y \\ \hline \Box_{lotr} \boxed{\begin{array}{c} x \\ \hline \text{Frodo}(x) \\ \text{struggle}(x) \end{array}} \\ \text{intriguing}(y) \end{array}}^{\text{W}} \quad \right\rangle$$
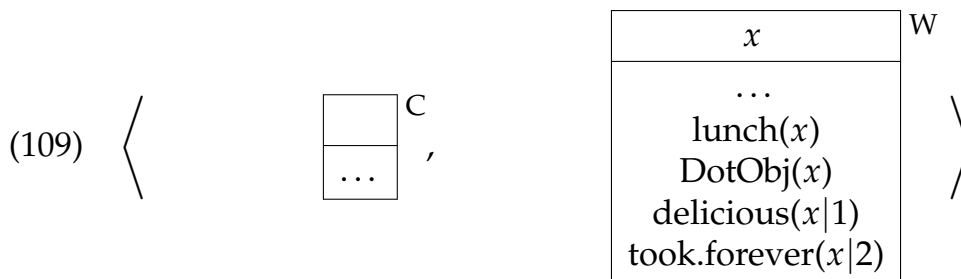
However, as has been explained in section 6.2.2, an ambiguity analysis seems to conflict with the anaphoric link in (67). In other words, it is incompatible

with what Recanati dubs the 'Anaphora-Coreference Principle' (i.e., if a pronoun is anaphoric on an antecedent name, the two terms co-refer (if they refer at all)) that is presupposed by previously discussed accounts of metafictional anaphora. However, Recanati argues, there are apparent counterexamples to this principle. Take the following sentence:

(108) Lunch$_i$ was delicious, but it$_i$ took forever. (Adapted from Asher 2011, p.11)

The pronoun 'it' is anaphoric on the noun 'lunch' of the preceding clause. However, 'lunch' and 'it' do not co-refer; 'lunch' refers to food (which was delicious) and 'it' refers to a social event (which took forever). Recanati argues that we can save the Anaphora-Coreference Principle by appealing to the notion of a dot-object (see e.g., Pustejovsky (1995); Luo (2012); Asher (2011)), i.e., "a complex entity involving several 'facets'" (Recanati, 2018, p.15). For instance, the noun 'lunch' is polysemous (i.e., it can refer to two (closely related) things: lunch qua food or lunch qua social event) and hence denotes a dot-object (represented as `food ● social event`) involving several facets (i.e., a food facet and a social event facet). Thus, in (108) 'lunch' and 'it' do actually co-refer (i.e to the dot-object 'lunch' or `food ● social event`), but the predicates 'being delicious' and 'taking forever' apply to different facets of the object (i.e., respectively to the food facet and to the social event facet). A workspace updated with (108) will thus look as follows, where the dot-object lunch (x) is predicated over through its food facet (represented as $x|1$, i.e., x predicated over through its first facet) and through its social event facet ($x|2$):

(109)
$$
\left\langle \quad \boxed{\begin{array}{c} \\ \hline \cdots \end{array}}^{\,C} , \quad \boxed{\begin{array}{c} x \\ \hline \cdots \\ \text{lunch}(x) \\ \text{DotObj}(x) \\ \text{delicious}(x|1) \\ \text{took.forever}(x|2) \end{array}}^{\,W} \quad \right\rangle
$$

According to Recanati, fictional names are also polysemous (i.e., they can refer to flesh and blood individuals or to abstract objects) and denote dot-objects (e.g., 'Frodo' denotes the dot-object `flesh and blood individual ● abstract object`). In metafictional statements 'Frodo' refers to this dot-object through its abstract object facet. In (explicit and implicit) parafictional state-

ments, 'Frodo' refers to this dot-object through its flesh and blood individual facet.

Recanati incorporates Zalta's distinction between encoding and exemplifying properties. The abstract object facet of Frodo both exemplifies properties such as being invented by Tolkien and encodes properties such as being a hobbit. Hence, the duality that is reflected in the two aspects of the dot-object Frodo, is also internal to the abstract object facet. Recanati agrees with Zalta that what properties the abstract object (facet) encodes is determined by our parafictional knowledge. However, whereas for Zalta the name 'Frodo' in parafictional discourse refers simply to an abstract object, for Recanati, it refers to a complex dot-object *through its flesh and blood facet*. Recanati can thus maintain that parafictional discourse that mentions Frodo is *primarily* about the flesh and blood hobbit.
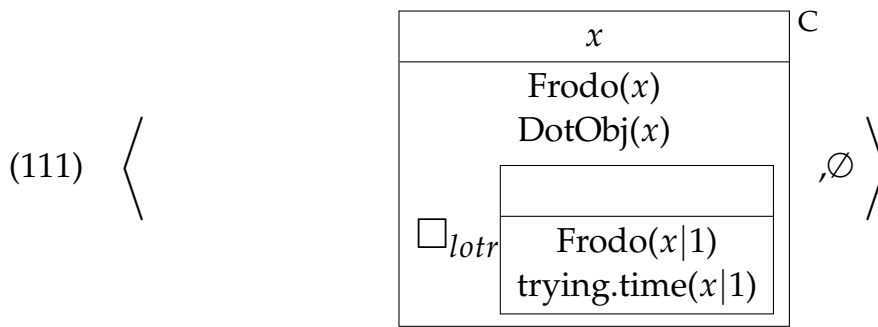
Fictive closure**

Applying Recanati's analysis to the workspace account suggests an adjustment of the fictive closure operation. I present a (simplified) representation of fictive closure** of fictional statement (25). First, we update the workspace with the content expressed by (25):

(110) $\left\langle \begin{array}{|c|} \hline \\ \hline \cdots \\ \hline \end{array}^{\text{C}} , \begin{array}{|c|} \hline x' \\ \hline \text{Frodo}(x') \\ \text{trying.time}(x') \\ \hline \end{array}^{\text{W}} \right\rangle$

At fictive closure** we update the common ground with discourse referents for dot-objects for any newly introduced fictional character and with parafictional information based on the content of the workspace. Dot-objects can be referred to as dot-objects (x), through their flesh and blood facet ($x|1$) (as is done in the parafictional condition) or through their abstract object facet ($x|2$). As in fictive closure* (see section 6.4.3), proper name conditions are doubled and also placed in the main box.
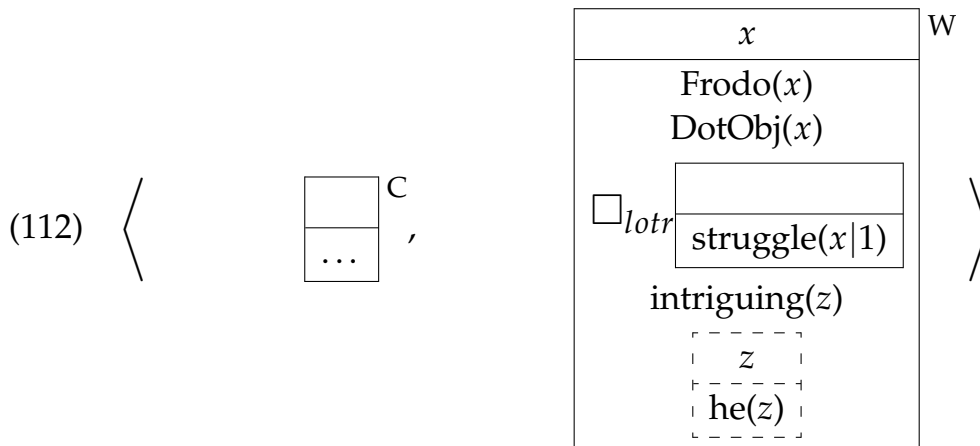
(111) $\left\langle \quad \begin{array}{|c|}\hline x \\\hline \text{Frodo}(x) \\ \text{DotObj}(x) \\ \Box_{lotr}\boxed{\begin{array}{c} \text{Frodo}(x|1) \\ \text{trying.time}(x|1)\end{array}} \\\hline \end{array}\!{}^{\text{C}} \;, \varnothing \quad\right\rangle$

As with fictive closure*, fictive closure** greatly complicates the basic (descriptivist) fictive closure operation by moving discourse referents for fictional entities outside of the DRS embedded under the fiction operator. Similar to fictive closure*, fictive closure** can be made into a more uniform mechanism if we combine a dot-object approach to fictional names with a de re version of the workspace account.

Addressing the challenge

A dot-object analysis of fictional characters solves the challenge posed by metafictional anaphora as in (67). A workspace updated with (67) looks as follows (before pronoun resolution):

(112) $\left\langle \quad \boxed{\begin{array}{c}\;\\ \hline \cdots \end{array}}^{\text{C}} \;, \quad \begin{array}{|c|}\hline x \\\hline \text{Frodo}(x) \\ \text{DotObj}(x) \\ \Box_{lotr}\boxed{\text{struggle}(x|1)} \\ \text{intriguing}(z) \\ \boxed{\begin{array}{c} z \\ \hline he(z)\end{array}} \\\hline\end{array}\!{}^{\text{W}} \quad\right\rangle$

As the DRS shows, the discourse referent x for the dot-object Frodo is accessible outside of the 'In *The Lord of the Rings*'-operator.[15] This dot-object

---

[15]Here I assume that it is common ground that there is a dot-object Frodo after the parafictional update in (67). This is not obvious. Arguably, in the case of (108), as long as we talk about lunch as food and there is no mention of lunch as a social event, we are really just talking about lunch as food. Only at the introduction of the zeugmatic discourse

can be predicated over through its abstract object facet by the metafictional condition, i.e., 'intriguing$(x|2)$'. Hence we can equate the discourse referents for 'Frodo' and 'he' and interpret the metafictional statement in (67).
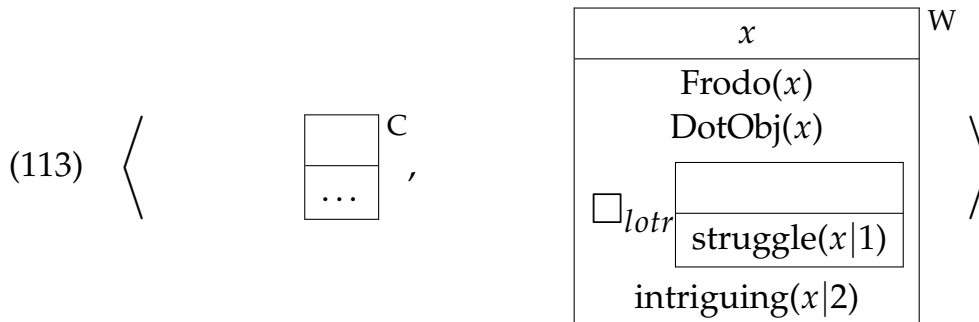
Although the solution is formally very similar to the solution offered by an abstract object account, it avoids a reoccurrence of the realist variant of the problem of the wrong kind of object while maintaining the Lewisian analysis of parafictional statements: the name 'Frodo' refers to the dot-object Frodo *through its flesh and blood facet* rather than to an abstract object in parafictional discourse. Moreover, like the object theoretic solution, a dot-object solution can easily be extended to account for parafictional anaphora and cases of co-predication across mixed metafictional and parafictional discourse. In general, dot-objects for fictional characters are always globally accessible in the main DRS (whether they have been introduced by fictional, parafictional or metafictional discourse). They can be predicated over through their flesh and blood facets in parafictional discourse and their abstract object facets in metafictional discourse. Whether the metafictional predication follows the parafictional predication, precedes it or occurs in the same statement is irrelevant.

Outlook

Although a fictive closure** variant of the workspace account that incorporates dot-objects seems to be a promising alternative to the simple descriptivist version of the workspace account (while still being consistent

---

does it become common ground that there is a dot-object (food ● social event) that we refer to. The same could be true about (67). However, as Recanati suggests, the (overt or covert) 'In fiction s'-prefix in parafictional discourse forces a 'metafictional perspective'; it makes us aware of the fictionality of the fictional characters and hence it is directly common ground at the parafictional update that we are referring to a dot-object including an abstract object facet.

with adhering to the Lewisian analysis of parafictional discourse), some of the details still need to be worked out. First, as Recanati himself also notes (2018, p. 43), the metaphysical status of dot-objects is controversial: is a dot-object simply an ordered tuple of its facets, e.g., $x = \langle x|1, x|2 \rangle$ (cf. Cooper (2007); Gotham (2017))? Must all facets of a dot-object exist in order for the dot-object to exist? These questions are especially pressing in the case of fiction where one of the facets of the dot-object (i.e., the flesh and blood individual facet) does not actually exist.

If the facets *do* all need to exist (in the actual world), this would mean that in order to engage in metafictional and parafictional discourse we have to assume the existence of both an abstract object *and* the existence of a flesh and blood creature. Arguably, this would make the dot-object x globally accessible through both facets. Such a strategy would entail adopting some kind of pretence analysis of parafictional discourse (cf. Evans (1982); Recanati (2018)) since we talk about fictional entities as if they really existed. In contrast, if we allow for facets to either really exist or (merely) exist in some fiction, we would only have to assume the existence of an abstract object and the existence *in some fictional worlds* of a flesh and blood creature in order to engage in metafictional and parafictional discourse. Arguably, this would imply that, although the discourse referent for the dot-object x is still always globally accessible through its abstract object facet, it is only accessible *through its flesh and blood facet* under the relevant fiction operator. This outcome fits neatly with an analysis of parafictional discourse as embedded assertions. Such a strategy would still amount to assuming some kind of realist/anti-realist ambiguity in fictional names, but would effectively stick to a descriptivist analysis of fictional names in parafictional discourse.

Alternatively, in order to avoid metaphysical assumptions about the existence of multifaceted dot-objects, Recanati suggests that the correct objects of study are in fact dot-*concepts* (i.e., concepts of dot-objects) or mental files (see Recanati (2012)) of dot-objects, rather than dot-objects.[16] Similarly, in the DRS's above, the discourse referents for dot-objects and their associated conditions can be understood as representing concepts rather than objects.

---

[16]Cf. Ninan's (2017) analysis of fictional names as referring to *possibilia*, (i.e., entities that exist at other possible worlds that the actual world) and Stokke's (forthcoming-a) recent proposal that fictional names refer to individual concepts.

6.5 Conclusions and further research

One of the central aims of this chapter has been to draw attention to the general challenge posed by metafictional discourse and especially the current desideratum for a theory of fictional names to account for anaphoric dependencies across parafictional and metafictional discourse while avoiding the problem of the wrong kind of object in its different variants. I have argued that the basic descriptivist version of the workspace account runs into difficulties with metafictional anaphora because the discourse referent for the fictional name introduced in the parafictional statement is not accessible outside of the 'In fiction s'-operator. This problem generalizes to other current dynamic approaches that involve separation of the content and discourse referents of the fictional narrative. I have evaluated four different accounts of metafictional anaphora: a descriptivist approach (that requires an additional account of how to accommodate the right kind of definite descriptions and cannot easily be extended to account for parafictional anaphora or co-predication across mixed metafictional/parafictional discourse), an approach based on changing the accessibility relations (that cannot straightforwardly be extended to a workspace account and runs into the anti-realist variant of the problem of the wrong kind of object), an abstract object account (that can give a consistent explanation of all anaphoric dependencies but forces us to give up the Lewisian analysis of parafictional discourse) and a dot-object account (that avoids the aforementioned problem but remains unclear on some crucial parts).

As mentioned in the introduction, any semantic account of fictional names will have to account for the use of pronominal anaphora across all acceptable types of mixed discourse. More specifically, it would be interesting to extend the described accounts of metafictional anaphora to other, potentially problematic, cases. For instance, an interesting type of discourse is what Bjurman Pautz (2008) dubs 'reports on fictional co-reference', i.e., anaphoric dependencies across belief reports on different people's parafictional beliefs as in (114). A related kind of discourse involves anaphoric dependencies across parafictional statements about different, inconsistent, fictional narratives. For instance, I can compare the location of Watson's war wound in two novels of the Sherlock Holmes series that are inconsistent in this respect as in (115). Or, suppose that apart from *The Lord of the Rings* Tolkien also wrote an alternative story (*The Lord of the Schmings*) in which the character

Gimli (a dwarf in *The Lord of the Rings*) is an elf. I could then felicitously say (116):

(114)   Bridget believes that Sherlock Holmes$_i$ is smart and Caroline believes that he$_i$ is smart. (Adapted from Bjurman Pautz, 2008)

(115)   In *A Study in Scarlet*, Watson's war wound$_i$ is on his shoulder but in *The Sign of Four*, it$_i$ is on his leg.

(116)   In *The Lord of the Rings*, Gimli$_i$ is a dwarf but in *The Lord of the Schmings*, he$_i$ is an elf.

Both types of discourse pose a challenge to a descriptivist version of the workspace account: discourse referents for Sherlock, Watson and Gimli are not accessible in the main box because they are embedded under a fiction operator (and in the case of (114) *also* embedded under a belief operator). Sentences (115) and (116) have the additional complication that, although the pronouns 'it' and 'he' are anaphoric on respectively 'Watson's war wound' and 'Gimli', it is not clear that these terms do in fact refer to the same fictional entity since they are ascribed inconsistent (individual-level) predicates in different narratives.

Discourses involving fictional co-reference or anaphoric dependencies across parafictional statements about distinct (inconsistent) narratives are reminiscent of both the phenomenon of counterfictional imagination (see e.g., Friend (2011b)) and Geach's Hob-Nob puzzle:

(117)   Hob thinks a witch$_i$ blighted Bob's mare, and Nob thinks she$_i$ killed Cob's sow. (Adapted from Geach, 1967)

Here the pronominal anaphora occur across two different propositional attitude reports and although the pronoun 'she' is anaphoric on 'a witch', there need not be one particular witch that is the object of thought of both Hob and Nob. Future research will have to determine how to account for discourses such as (114), (115) and (116) and specify their relation to other puzzles.

As Bjurman Pautz notes, a possible strategy to deal with fictional co-reference as in (114) is to adopt an object theoretic approach so that the terms 'Sherlock Holmes' and 'he' in (114) refer to a globally accessible abstract object. Arguably, a dot-object approach would function similarly. Moreover, both strategies can be extended to account for discourse such as (115) and (116). Concerning the ascription of inconsistent properties to the same fictional character we can maintain, in an object theoretic account,

that 'Gimli' and 'he' in (116) refer to the same *consistent* fictional character that only encodes being a dwarf (because Gimli is native to *The Lord of the Rings*). In cases where there is no clearly authoritative narrative (e.g., in (115)), we can analyse 'Watson's war wound' and 'it' as referring to the same fictional character that encodes *inconsistent* properties (assuming a theory of fictional truth that allows for inconsistent explicit parafictional statements to be both true). On a dot-object approach, we can – assuming we do not allow for inconsistent 'flesh and blood facets' – analyse the terms 'Watson's war wound' and 'it' in (115) and 'Gimli' and 'he' in (116) as referring to the same dot-objects but through different 'flesh and blood facets'. We thus require dot-objects with three or four different facets: In the case of (116) a 'flesh and blood dwarf' facet, a 'flesh and blood elf' facet, and one or two abstract object facets (depending on whether we allow for inconsistent abstract objects).

Appendix

Examples of all six possible types of mixed discourse with fictional, parafictional and metafictional statements.

Type of mixed discourse	Example
fictional/ parafictional	Hans$_i$ and Gretel$_j$ approached the skyscraper. "Maybe you should have a look inside, Gretel. They might have candy", whispered Hans$_i$.* Gretel$_j$ moved closer... *In this story, she$_j$ is the hero that saves the day. He$_i$ is the villain. (Semeijn and Zalta, 2021, p.7)
fictional/ metafictional	In order to capture the witch, Mary$_i$ travelled to the woods and disguised herself as a potato.* In the woods she$_i$ encountered many perils... *I know this is weird but I invented her$_i$ while eating chips. (Semeijn and Zalta, 2021, p.6-7).
parafictional/ fictional	In the story I made up yesterday, a wizard called Brian$_i$ falls in love with a cauldron. Let me tell it to you: One day, he$_i$ was alone in his$_i$ study trying out a new love-potion recipe... (Semeijn and Zalta, 2021, p.6)
parafictional/ metafictional	In *The Lord of the Rings*, Frodo$_i$ goes through an immense mental struggle to save his$_i$ friends. Ah yes, he$_i$ is an intriguing fictional character!
metafictional/ fictional	Frey$_i$ is a fictional character I made up and is the protagonist of my newest story. Here it is: One day she$_i$ was walking through the woods near her home... (Semeijn and Zalta, 2021, p.6)
metafictional/ parafictional	Sherlock Holmes$_i$ is a fictional character created by Conan Doyle. In Conan Doyle's stories, he$_i$ is a private detective who investigates cases for a variety of clients, including Scotland Yard. (Adapted from Recanati, 2018, p.37)

7 The 'In' and 'According to' operators

This chapter is a rewritten version of 'The 'In' and 'According to' operators' in *Proceedings of the ESSLLI & WeSSLLI Student Session 2020*. The most substantial differences between this chapter and the proceedings paper include: First, a removal of the discussion on the Lewisian analysis of "In s, ϕ". Second, the addition of a further research section 7.4.2 which includes a brief discussion of how the proposed analysis of "According to s, ϕ" relates to the previous discussion of fictive closure and bald-faced lies.

7.1 Introduction

In chapter 3 I analysed parafictional discourse (i.e., discourse about the content of some fiction) as being covertly or overtly of the form 'In/According to s, ϕ'. In other words, parafictional statements contain either an 'In s' operator (henceforth abbreviated as **In**) or an 'According to s' operator (henceforth abbreviated as **Acc**):

(118) a. In *The Hobbit*, Bilbo travels to the Lonely Mountain.
       b. According to *The Hobbit*, Bilbo travels to the Lonely Mountain.

Following common practice, I have so far largely ignored any semantic difference between the operators **In** and **Acc**. However, the astute reader may have noticed that the Lewisian analysis of parafictional discourse that I adopt has been formulated (following Lewis) for statements of the form 'In s, ϕ' rather than 'In/According to s, ϕ' (see section 4.2.2). The reason for this is that in the current chapter I suggest, contra common practice, that there is in fact a relevant semantic difference between **In** and **Acc** and suggest that **In** is the primary fiction operator.

In providing analyses of parafictional discourse almost[1] all philosophers (e.g., Zucchi (forthcoming); Recanati (2018); Zalta (1983)) and semanticists (e.g., von

[1]Notable exceptions are Sainsbury (2014) and Voltolini (2019) who argue that **In** involves a more 'distanced' stance towards the fiction than **Acc**. However, the authors do not

Fintel and Heim (2011)) treat **In** and **Acc** on a par, i.e., (118a) and (118b) receive the same truth conditions. One of the main objectives of this chapter is to establish that there are in fact interesting semantic differences between **In** and **Acc**. These differences have probably remained largely unrecognized or glossed over because semanticists of fiction traditionally focus on providing analyses for reports on the content of *fictional* media (i.e., parafictional statements) only; since **In** and **Acc** both seem acceptable in such statements (e.g., in (118a) and (118b)), a uniform semantic analysis seems justified. To tease apart **In** and **Acc** I adopt a broader perspective in this chapter and consider reports on the content of media whether fictional or non-fictional, i.e., so-called 'contensive' (Ross, 2012) or 'paratextual' (Zucchi, 2001) statements. For instance, apart from parafictional statements (118a) and (118b), the following report on the content of Monk's biography *Ludwig Wittgenstein: The Duty of Genius* is also a contensive statement:

(119)   According to *Ludwig Wittgenstein: The Duty of Genius*, Wittgenstein worked as a hospital porter during WWII and advised patients not to take the drugs they were prescribed.

Moreover, the observations and analyses concerning the use of **In** and **Acc** in contensive statements that are discussed in this chapter apply not only to contensive statements about written or spoken narratives (e.g., *The Lord of the Rings* and *Ludwig Wittgenstein: The Duty of Genius*) but also to contensive statements about non-verbal media such as movies and pictures. The current discussion thus extends to for instance statements about the content of the *Star Wars* saga:

(120)   a.   In the *Star Wars* saga, Darth Vader is a Sith Lord.
        b.   According to the *Star Wars* saga, Darth Vader is a Sith Lord.

Hence this chapter also adopts a broader perspective compared to the rest of the dissertation (where I focus on verbal narratives) in this respect.

---

connect their semantic accounts of the operators to existing linguistic work on **Acc** as is done in this chapter. Nor do they discuss the linguistic observations concerning the diverging behaviour of **In** and **Acc** introduced in this chapter. Rather, part of their debate is on whether the following minimal pair illustrates the semantic difference between **In** and **Acc**:

   (v)   In *War and Peace*, there are both fictional and real characters.

   (vi)   According to *War and Peace*, there are both fictional and real characters. (Sainsbury, 2014, p.278)

Whereas Sainsbury takes (v) to be true and (vi) false, Voltolini takes both to be false. I do not further discuss these types of statements in this chapter since my focus lies on the use of **In** and **Acc** in parafictional statements and both (v) and (vi) seem to have a distinct metafictional flavour, i.e., talk about fictional entities *as fictional entities*. See chapter 6 for a more elaborate discussion of metafictional discourse.

As discussed in chapter 4, I assume the widely adopted Lewisian (1978) possible world analysis for **In**. Roughly: "**In** *s*, *ϕ*" is true iff in worlds compatible with *s*, *ϕ*. In what follows I will first discuss the proposed analysis of **Acc** (section 7.2). In line with Krawczyk's (2012) analysis of 'According to *s*', contensive statements with **Acc** are analysed as indirect speech reports. Roughly: "**Acc** *s*, *ϕ*" is true iff *s* asserts that *ϕ*. Second, I will explore three clusters of novel observations concerning the divergent linguistic behaviour of **In** and **Acc** that a uniform treatment of the operators *cannot* but that the proposed semantic analyses *can* explain. These observations add to existing observations in recent linguistic literature that show that there is a crucial difference between **Acc** and other intensional operators (e.g., Krawczyk (2012); Kaufmann and Kaufmann (2020); Bary and Maier (2020)). The novel observations relate to the fictionality of the medium that is reported on (section 7.3.1), reporting explicit and implicit content (section 7.3.2) and tense use in contensive statements (section 7.3.3).

7.2 Semantic analysis of the 'According to *s*'-operator

Contensive statements that feature the operator **Acc** are analysed as a type of indirect speech report, i.e., reports on what a medium asserts or asserted (but not on what it 'will assert'[2]):

"According to *s*, *ϕ*" is true iff *s* asserts/asserted that *ϕ*

This analysis of contensive statements with **Acc** is in line with Krawczyk's (2012) and Kaufmann and Kaufmann's (2020) analysis of the general (i.e., also outside of contensive statements) use of the phrase 'According to *s*'. These semanticists treat **Acc** not as a simple intensional operator (cf. von Fintel and Heim (2011)) but rather treat statements with this phrase as indirect speech reports. Indeed, such an analysis fits the use that **Acc**, unlike **In**, has outside of contensive statements; **Acc**

---

[2]Although the speech verb in speech reports can be future tensed when reporting on (or predicting) future speech events, it seems that **Acc** only has interpretations where it is used to report on either past or ongoing speech events. If I know Joe doesn't like seagulls (because he always angrily throws stones at them) and I know that he is going to assert that seagulls are the worst (because he raised his hand at a 'seagull-lovers'-seminar) but I have never actually heard John *assert* that seagulls are the worst, I can felicitously say (ix) but not (x):

(ix)   Joe will assert that seagulls are the worst.

(x)   # According to Joe, seagulls are the worst.

can be used to report not only on the content of a medium but also on what some *person* asserted:

(121)   a.   According to Joe, seagulls are the worst.
        b.   # In Joe, seagulls are the worst.

As Anand and Korotkova (2019) note, such reports behave like regular indirect speech reports. For instance, whereas belief reports can be followed by a denial of the embedded content having been said, speech reports cannot:

(122)   a.   Joe thinks that seagulls are the worst. He never said that, though.
        b.   # Joe asserted that seagulls are the worst. He never said that, though.

Likewise, it seems that (121a) cannot be followed by a denial of the embedded content having been said:

(123)   ? According to Joe, seagulls are the worst. He never said that, though.

Reports with **Acc** behave like *indirect* speech reports in that the report does not have to repeat the exact phrasing of the reported speech. Suppose Joe's exact words were: "I hate seagulls! There is no animal that is worse". As in indirect speech report report (124b) (and unlike in direct speech report (124a)), we can report on this speech act by combining **Acc** with the paraphrase "seagulls are the worst":

(124)   a.   # Joe said: "Seagulls are the worst".
        b.   Joe said that seagulls are the worst.
        c.   According to Joe, seagulls are the worst.

    Anand and Korotkova (2019) argue that this analysis of **Acc** can not only apply to reports on what some speaker has said but also to reports on what some inanimate object has 'said' as long as the object is a repository of propositional information (or 'ROI subject', see Anand and Hacquard (2014)) such as books, theories, films or lecture notes. In other words, we can report on the content of a medium by talking about the medium as the 'agent' of a communicative act. For instance, we can talk about what a book 'tells us'.[3] Hence **Acc** can feature in contensive statements which as a result are interpreted as reports on what some medium (e.g., *The Lord of the Rings* or a news report) asserts – rather than reports on what the author of the medium asserts.[4]

---

[3]I assume that nonverbal media (e.g., the *Star Wars* saga) are also ROI subjects and hence we also report on those as 'telling us' things. In case the reader thinks nonverbal media don't assert in this way, they may read 'the *Star Wars* saga' as 'the script of the *Star Wars* saga'.

[4]This semantic analysis is akin to Zalta's (1987) analysis of parafictional statements in general (with **In** or **Acc**) as reporting on what a fictional narrative asserts. Such analyses raise some questions concerning the notion of 'assertion' at play here. For instance, a

Before moving on, it is instructive to highlight two features of the speech act of assertion that will be relevant later. First, since assertions are non-fictional statements, when *a* asserts *φ* this means that *a* states that *φ* is true in the *actual world*, i.e., *a* communicates that the actual world is in the set of *φ* worlds. Likewise, when some medium is reported on as making an assertion, this means that it is treated as stating something about the actual world. In other words, it is reported on as if it is non-fiction.[5] It is possible to formulate the semantic analysis of **Acc** with speech verbs that are similar in meaning such as "say" or "express". I use "assert" because, as will become clear later (section 7.3.1), I want to restrict the analysis to reports on speech acts that are clearly commitment inducing. Possibly, "say" or "express" are too generic (e.g., fictional or presuppositional contents may be *said* or *expressed* but are not *asserted*).

Second, unlike simple intensional operators, indirect speech reports are generally not closed under logical entailment (see e.g., von Stechow and Zimmermann (2005); Sæbø (2013); Maier (2019)). Consider the following belief report:

(125)   Anne believes that Chrissy is cool.

Under a simple modal analysis of belief (Cf. Hintikka (1962)), (125) is true iff Chrissy is cool in all possible worlds that are compatible with Anne's beliefs. Given that Chrissy being cool implies (amongst other things) that there is at least one cool person, it is also true in all possible worlds that are compatible with Anne's beliefs that there is at least one cool person. Hence, it follows from (125) that Anne also believes that there is at least one cool person. Moreover, logical truths (e.g., that bachelors are unmarried men) are necessary, i.e., they are true in all possible worlds. Hence it is also true in all worlds compatible with Anne's beliefs that bachelors are unmarried men; Anne also believes this. In short, under an intensional analysis of belief, any agent that forms beliefs is logically omniscient, i.e., believes all logical truths and all logical consequences of their beliefs.

Although such an intensional analysis may be fine when we offer a description of an idealized rational agent's beliefs, it cannot straightforwardly be applied to speech reporting. Consider the following variants of speech reports:

(126)   Anne asserts/says/claims/yells/mutters/whispers that Chrissy is cool.

On a simple intensional analysis of asserts/says/claims/yells/mutters/whispers, (126) is true iff Chrissy is in fact cool in all possible worlds that are compatible with what Anne asserts/says/claims/yells/mutters/whispers. Like the intensional analysis of belief above, the intensional analysis here thus implies that (126) entails that

---

book cannot have any beliefs and hence a book's assertion cannot be construed as a proposal to update the common ground between speaker and hearer if it is construed as common belief or common acceptance. I will ignore these issues for now.

[5]Cf. Murday (2010) who argues that use of fiction operators such as **Acc** in parafictional statements relates the content of the fictional narrative to the actual world.

Anne also asserts/says/claims/yells/mutters/whispers all logical consequences of Chrissy being cool (e.g., that there is at least one cool person) *and* all logical truths (e.g., that bachelors are unmarried men). But does it follow from (126) that Anne asserts/says/claims/yells/mutters/whispers that there is at least one cool person? Maier (2019) argues that for many so-called 'descriptive communication verbs' (e.g., yells/mutters/whispers) the entailment is definitely off and hence the intensional analysis cannot work. For less descriptive verbs (e.g., say/assert/claim) the entailment will sometimes seem acceptable. For the latter type of verbs we can follow von Stechow and Zimmerman's (2005) suggestion to analyse indirect speech reports with 'say' as ambiguous between a strict reading – where they are *not* closed under entailment – and a non-strict reading – where they *are* closed under entailment.

However, even on a non-strict reading of speech verbs such as 'say', 'assert' and 'claim', a simple intensional analysis will still not suffice: Although we may accept that we can derive from (126) that Anne asserted that there is at least one cool person, in any case it still does not follow that Anne asserted that bachelors are unmarried men. Moreover, even on a non-strict reading, speech reports are only closed under entailment *to a certain extent*, i.e., not *all* logical entailments of what was stated are reportable with indirect discourse. For instance, it does not follow from (126) that Anne asserted that Chrissy is cool or a murderous clown (even though Chrissy being cool or a murderous clown *does* strictly speaking follow from Chrissy being cool). In other words, even on a non-strict reading of 'say', 'assert' and 'claim', only a subset of the entailments of what was explicitly stated are also actually 'said', 'asserted' and 'claimed'.[6]

In the above semantic analysis of **Acc** "asserts" is to be read non-strictly, i.e., *s* asserts that ϕ iff *s* explicitly states ϕ or ϕ is properly entailed by what *s* explicitly states.

7.3 The diverging behaviour of 'In' and 'According to'

Now that I have presented my semantic analyses of **In** and **Acc**, I turn to three linguistic observations concerning the diverging linguistic behaviour of **In** and **Acc** (and some qualifications to them). Current analyses of contensive statements do not distinguish **In** from **Acc** and therefore do not explain these observations. I will argue that the Lewisian analysis of **In** and the above analysis of **Acc** can account for them.

---

[6]See e.g., Brasoveanu and Farkas (2007); Sæbø (2013); Bary and Maier (2020); Abreu Zavaleta (2019) for some further discussion on this topic.

7.3.1 Fiction/non-fiction

Observations

A central observation concerning **In** and **Acc** is that whereas contensive statements about fiction can be formulated with both **In** and **Acc**, contensive statements about non-fiction with **In** rather than **Acc** are typically unacceptable. Consider the following minimal pairs of statements:

(120) a. In the *Star Wars* saga, Darth Vader is a Sith Lord.
b. ? According to the *Star Wars* saga, Darth Vader is a Sith Lord.

(127) a. # In *Ludwig Wittgenstein: The Duty of Genius*, Wittgenstein worked as a hospital porter during WWII and advised patients not to take the drugs they were prescribed.
b. According to *Ludwig Wittgenstein: The Duty of Genius*, Wittgenstein worked as a hospital porter during WWII and advised patients not to take the drugs they were prescribed.

Whereas use of **Acc** seems appropriate to report on the content of fictional and non-fictional media, use of **In** seems restricted to reports on the content of fictional media. Even stronger, this fiction/non-fiction preference is also mirrored in our use of **Acc**. As noted before, **Acc** can be (and is) used to report on the content of fictional media. I therefore generally do not mark such uses of **Acc** as infelicitous. However, use of **In** does typically sound more appropriate in parafictional statements than use of **Acc**, e.g., (120a) and (120b) are both acceptable but (120a) is a more natural way of talking about the content of the *Star Wars* films. Thus the general picture that is sketched is that the canonical use of the operators links **In** to fiction and **Acc** to non-fiction.

The observation made above can be qualified in several ways. First, use of **Acc** in contensive statements about fiction is not always unnatural and sometimes is even more appropriate than use of **In**. Contensive statements that report on the content of a fictional medium that is viable for 'export'[7] (i.e., content that we may take to be not only true in the fiction but also true about the actual world) display such preferences. Such exported content can consist of empirical facts that were explicitly stated in a medium. For instance, I may read the following in Fleming's novel *Thunderball*: "New Providence, the island containing Nassau, the capital of the Bahamas, is a drab sandy slab of land fringed with some of the most beautiful beaches in the world", and learn from this that (actually) Nassau is the capital of the Bahamas. Alternatively, we can export general truths that follow from what was explicitly stated or shown in a fiction. For instance, I may learn that (actually) it is never too late to redeem yourself from watching the *Star Wars* saga even though

---

[7]See section 4.6.2 for a more elaborate discussion of export of fictional truth.

this is never stated explicitly in the medium. Especially in contensive statements that report on general fictional truths that are viable for export we can observe an increased acceptability of **Acc**. For instance, consider the following:

(128)   a.   In the *Star Wars* saga, it is never too late to redeem yourself.
         b.   According to the *Star Wars* saga, it is never too late to redeem yourself.

Use of **Acc** is decidedly more natural in (128b) than in contensive statements that report on fictional content that is *not* viable for export such as (120b). Arguably there is even a small preference for use of **Acc** as in (128b) over use of **In** as in (128a). This latter intuition become even stronger when we consider 'fictions' whose point is clearly to teach us something about the actual world. Consider the following contensive statements about Searle's Chinese room thought experiment:

(129)   a.   According to the Chinese room thought experiment, something that manipulates symbols based on syntax alone, does not truly understand a language.
         b.   ? In the Chinese room thought experiment, something that manipulates symbols based on syntax alone, does not truly understand a language.

Here use of **Acc** to report on the fictional content that is viable for export in (129a) is appropriate whereas use of **In** in (129b) is unnatural.

Second, use of **In** is not in fact unequivocally wrong for contensive statements about non-fictional media. Zucchi provides the following example of a contensive statement featuring **In** about Woodward's biography *Shadow*:

(130)   a.   In *Shadow*, Clinton only cares about sex and golf. (Zucchi, 2001, p.350)
         b.   According to *Shadow*, Clinton only cares about sex and golf.

Not only use of **Acc** but also use of **In** is acceptable in this non-fiction contensive statement. However, note that such use of **In** is restricted to reports on subjective viewpoints or portrayals that are expressed by some medium rather than objective facts. Use of **In** here seems to signal distancing from the reported content. Likewise, a contensive statement with **In** that reports on an objective fact expressed by *Shadow* sounds as odd as (127a):

(131)   a.   # In *Shadow*, Clinton was born in Arkansas.
         b.   According to *Shadow*, Clinton was born in Arkansas.

Analysis

The provided analyses account for the above observations. First, when we report on the content of some non-fictional source *s* (e.g., a biography, news report or encyclopedia entry), we will report on the medium as telling us (or asserting) something about the actual world – not as some story that is compatible with some

set of worlds that may or may not include the actual world. Hence we have a strong preference for **Acc** in contensive statements about non-fiction. By contrast, when talking about the content of a fictional medium *s* it is appropriate to consider what is true in the set of *s* worlds without reporting on *s* as asserting anything about the actual world. Hence **In** is appropriate whereas use of **Acc** (i.e., reporting on the content of a fiction story as if it relates to the actual world) is less natural. Thus there is a general preference to use **In** for reports on fiction and to use **Acc** for reports on non-fiction.

As I have shown, however, although there may be a preference for **In**, **Acc** is in fact generally acceptable for contensive statements about fiction (e.g., (118b), (120b)). The semantic analysis of **Acc** suggests that this is because it is considered generally admissible to report on the content of a fictional medium by talking about it as something that asserts something about the actual world.[8] Usually, such use of **Acc** will sound unnatural because fictional media are standardly not considered to be appropriate authorities or sources for claims about the actual world. Hence a statement such as (120b), which means something like "The *Star Wars* saga asserts that Darth Vader is a Sith Lord", sounds awkward. In fact, it is strictly speaking not even true![9] The *Star Wars* saga, being a work of fiction, does not really assert anything about the history of the galaxy. We merely pretend that it does when engaging with the fiction. Similarly, the analysis suggests that use of **Acc** in parafictional discourse such as (118b) is proper when we report on the content of fictions by engaging in an extension of this original pretence, i.e., talking about the fictional medium as non-fictional (cf. Evans (1982); Recanati (2018)). On the current analysis, parafictional statements with **Acc** such as (118b) are thus an interesting hybrid type of discourse: They are parafictional because they are reports on the content of a fictional medium *s*, but also constitute an unofficial extension of the original fictional discourse of *s* (which makes them strictly speaking false).

---

[8]Hence the analysis provided seems to be in line with Friend's (2017) claim that *all* fictional narratives are essentially to be interpreted as being about the actual world, even when the described events take place in an outlandish magical realm where for instance Earth does not even exist.

[9]In case the reader finds this counterintuitive, we can opt for a semantic analysis of **Acc** in terms of "expressing":

"According to *s*, ϕ" is true iff *s* expresses/expressed that ϕ

Such an analysis would deem contensive statements such as (120b) true (cf. Voltolini (2019); Murday (2010)); fictional media may not 'assert' (all) their content but they do 'express' it. However, the above analysis is unable to account for any of the observations described in the current section concerning preferences for **In** and **Acc** to report on fictional and non-fictional media. Hence I opt for an analysis of **Acc** that can account for all three observations described in this chapter at the cost of making statements such as (120b) strictly speaking false.

Parafictional statements with **In** such as (120a), however, involve no such pretence and can be really true.

The analysis also explains why contensive statements that report on content viable for export constitute exceptions to the general awkwardness of the use of **Acc** for fiction. Independently from what analysis we adopt of export[10], it is generally assumed that content viable for export is somehow licensed by the author to be taken as *actually* true (in addition to fictionally true). We *do* in fact take the medium to tell us something about the actual world. Hence the proposed analysis of **Acc** predicts a higher acceptability rate for **Acc** when reporting on fictional content that is viable for export. For instance, although the *Star Wars* saga does not tell us anything about the history of the galaxy, we could take it to be a proper source of (non-fictional) moral truths. Hence, although we take (120b) to be strictly speaking false, we take (128b) to be true (i.e., that the *Star Wars* saga asserts that it is never too late to redeem yourself). This higher acceptability may even trump the acceptability of **In** when it obviously is the point of the fiction to tell us something about the actual world (as is the case of thought experiments such as Searle's *Chinese room*).

The analyses also account for the fact that sometimes **In** may be appropriate for contensive statements about non-fiction as in (130a). According to our semantic analysis of **In**, (130a) roughly means that in the worlds compatible with *Shadow*, Clinton only cares about sex and golf. In other words, the medium is not presented as telling us something about the actual world. Rather, because we are reporting on subjective content it is acceptable to report on what the worlds compatible with the medium are like (i.e., report on *Shadow* as if it is fiction). The perceived distancing from the reported content by the speaker of the contensive statement seems to be the result of pragmatic implication (i.e., given that the relevant medium is non-fictional, why doesn't the speaker report on its content as asserting something about the actual world?)

7.3.2 Explicit/implicit content

Observations

The second observation about the difference between **In** and **Acc** relates to whether the reported content is explicit or implicit in the medium. Semanticists of fiction often assume some version of Lewis' (1978) Reality Principle: we assume the fictional worlds to be as much like the actual world as the story permits. In other words, we can distinguish two types of fictional truths: 'Explicit fictional truth', i.e., propositions that are explicitly stated in a story (or follow directly from what was explicitly stated) and 'implicit fictional truth', i.e., propositions that are assumed to be fictionally true because we consider them to be actually true and the story has

---

[10]See section 4.6.2.

not forced us to revoke them. For instance, it is explicitly fictionally true in *The Lord of the Rings* that Frodo inherits Bag End because this follows directly from some of the statements in the novels. On the other hand, it is implicitly fictionally true in *The Lord of the Rings* that water is H_2O because we believe this to be actually true and nothing in the novels contradicts this information.

Semanticists of fiction generally allow for both implicit and explicit fictional truths to feature in parafictional statements. This type of approach ignores important differences in linguistic behaviour between **In** and **Acc**. **In** is appropriately used to report on both implicit and explicit fictional truth. Consider the following statements:

(132)   In *The Lord of the Rings*, Frodo inherits Bag End.

(133)   In *The Lord of the Rings*, water is H_2O.

Acc displays different behaviour. To the extent that use of **Acc** to report on fictional content is acceptable at all, **Acc** can only appropriately be used to report on explicit fictional truth. Consider the following statements:

(134)   According to *The Lord of the Rings*, Frodo inherited Bag End.

(135)   # According to *The Lord of the Rings*, water is H_2O.

Use of **Acc** is thus restricted to parafictional statements that report content that is explicitly stated in the medium or follows directly from what was stated.

This observation generalizes to contensive statements about non-fiction. Consider the following contensive statements about a news report that reports on a drought (but does not state anything about the molecular structure of water):

(136)   According to this news report, there was a terrible drought.

(137)   # According to this news report, water is H_2O.

Although the fact that water is H_2O may be assumed to be true (by speaker and hearer alike) when engaging with this news report, such 'implicit truths' cannot feature in contensive statements with **Acc**. Again, **Acc** is only appropriate to report on what was explicitly stated in the medium or what follows directly from this.

Analysis

The proposed analyses can account for the above observations concerning implicit and explicit content. First, the Lewisian analysis of **In** was formulated so as to include implicit fictional truths. The worlds compatible with *s* are the worlds where *s* is told as known fact that are *as similar as possible* to our conception of the actual world. In other words, everything that we believe to be actually true will be true in the worlds compatible with *s* unless *s* contradicts it. So even though the fact that water is H_2O is never stated explicitly (nor follows from anything that was

stated) in *The Lord of the Rings*, still it is true in the worlds compatible with *The Lord of the Rings* because the worlds where *The Lord of the Rings* is told as known fact that are *closest* to our conception of the actual world are worlds in which water is H_2O. Thus **In** can appropriately be used to report on such implicit content.

Second, the analysis of contensive statements with **Acc** as indirect speech reports excludes reports on implicit content. Remember that under the non-strict reading that we adopt of "asserts" in the semantic analysis of **Acc**, *s* asserts only those things that are explicitly stated by *s* and some of the entailments of what *s* explicitly stated. Information that is merely assumed by *s* but that is neither said nor even entailed by what was said cannot feature in indirect speech reports (e.g., from the fact that Anne asserts that Chrissy is cool we cannot derive that Anne asserts that Chrissy plays basketball even though it may be common ground that she does). Likewise, it is not appropriate to report on 'content' that was not stated explicitly (or follows from what was stated) in some medium (e.g., *The Lord of the Rings* or a news report on a drought) with **Acc** even though this information may arguably be part of what is assumed to be true by the medium.[11]

7.3.3 Tense use

Observations

The third and last observation concerning **In** and **Acc** that I will discuss relates to tense use preferences in contensive statements. As has been observed by Zucchi (2001), parafictional statements with **In** display a preference for present tense use while past tense, although often acceptable, sounds awkward and future tense simply sounds wrong.[12] Parafictional statements with **In** trigger this preference for present tense independently from whether the embedded content includes an eventive or stative verb. Consider for example the following contensive statements about the Harry Potter novels:

(138)   In the Harry Potter novels, there **are/?were/#will be** wizards in England.

(139)   In the Harry Potter novels, Snape **kills/?killed/#will kill** Dumbledore.

Whereas (138) includes a stative verb and (139) contains an eventive verb, both contensive statements trigger a preference for present tense.

---

[11]This semantic difference between **In** and **Acc** suggests that the proper parafictional test case sentences (i.e., appropriate to check our intuitions against about fictional truth) should be formulated with **In** rather than **Acc**.

[12]The prohibition against past and future tense in parafictional statements is not absolute. Consider: "In Patrick O'Brian's first novel, Jack Aubrey was a post captain, in his new novel, he is a commodore, in the next novel he will be an admiral." (Zucchi, 2001, p.334).

Moreover, parafictional statements with **In** display a preference for present tense independently from when the events described in the fiction supposedly take place. Consider for example the following contensive statements about the Harry Potter novels, the *Star Wars* saga and the *Star Trek* series for which the time of the relevant fictional events and states described respectively overlap, precede and succeed the fictional counterpart of the utterance time of the contensive statement:

(138) In the Harry Potter novels, there **are/?were/#will be** wizards in England.

(140) In the *Star Wars* saga, Luke **destroys/?destroyed/#will destroy** the Death Star.

(141) In the *Star Trek* series, Earth **colonizes/?colonized/#will colonize** Mars in the year 2103.

This preference for present tense does not generalize to parafictional statements with **Acc**. Rather, to the extent that **Acc** is at all acceptable to report on fictional content, preferences for tense use within these statements seems to depend on the time of the events described in the narrative relative to the utterance time of the contensive statement, i.e., whether, at the time of utterance, the relevant fictional events took, take or will take place:

(142) According to the Harry Potter novels, there **are/#were/#will be** wizards in England. (*stative/overlap*)

(143) According to the *Star Wars* saga, Luke **#destroys/destroyed/#will destroy** the Death Star. (*eventive/precede*)

(144) According to the *Star Trek* series, Earth **#colonizes/#colonized/will colonize** Mars in the year 2103. (*eventive/succeed*)

In fact, this is true for contensive statements with **Acc** in general, i.e., tense use in contensive statements with **Acc** about non-fictional media also seems to depend on the time of the events described in the medium relative to the utterance time of the contensive statement. Consider tense use in the following statements about the content of news reports that report on respectively protests going on at this moment, a robbery last night and tomorrow's weather:

(145) According to this news report, there **are/#were/#will be** protests in Amsterdam.

(146) According to this news report, masked men **#rob/robbed/#will rob** the Regio Bank in Erp.

(147) According to this weather forecast, it **#is/#was/will be** extremely dry.

169

Analysis

The proposed semantic analyses of **In** and **Acc** can account for these observations. First, the analysis of **In** predicts a preference for present tense in contensive statements. To see why, let's first consider tense use under other intensional operators such as *believe*:

(148)   Adeela believes that Sara was nervous.

Because this propositional attitude report is a report of a current belief (i.e., the attitude verb is in present tense), the tense use in the embedded clause tells us whether Adeela believes Sara to be nervous before, during or after the time of utterance of (148).[13] In the above example: if (148) is uttered at t_1 then (148) is true iff in worlds compatible with what Adeela believes at t_1, Sara *was* nervous at t_1 (i.e., *is* nervous at some t where $t < t_1$).

   In, although also an intensional operator, functions somewhat differently. Whereas someone's beliefs may change over time (e.g., Adeela might change her mind about whether Sara is in fact nervous), the content of a story or medium (e.g., the Harry Potter novels) consists in an abstract set of statements or system of axioms that is timeless. The Harry Potter story today is not going to differ from the Harry Potter story tomorrow; it is eternally the same abstract object. Hence, although we report on what some agent's beliefs are at a certain point in time in (148), in contensive statements we do not report on what the Harry Potter novels are like at a certain point in time. Reconsider the present tense version of (138):

(138)   In the Harry Potter novels, there are wizards in England.

Even though (138) is uttered at a specific point in time t_1, (138) does not mean that in worlds compatible with the Harry Potter novels at t_1, wizards are in England at t_1. Rather, (138) uttered at t_1 is true iff in worlds compatible with the Harry Potter novels (simpliciter), there are wizards in England (at some t). Hence, because it is true that there are wizards in England at a specific point on the fictional timeline of the Harry Potter novels, (138) is true. Indeed, given this fact, the past and future tense versions of (138) (although they sound odd or infelicitous) are also strictly speaking true on this analysis. It is true on some point in the timeline of the Harry Potter worlds that there *were* wizards in England and similarly there is such a point where there *will be* wizards in England. In other words, the Lewisian analysis of **In** (since it does not designate a specific time of evaluation) strictly speaking permits present, past and future tense use in contensive statements:[14]

---

[13]Reports with past or future tense attitude verbs (e.g., 'Adeela believed/will believe that Sara is nervous') pose additional complications since tense in these reports can be bound rather than indexical (see Abusch (1997); Toshyuki and Sharvit (2012)).

[14]In fact, Lewis (1978), uses both past tense and present tense in his examples of parafictional truths.

(149) In the Harry Potter novels, there ?were/#will be wizards in England.

I suggest that this fact licenses a gnomic or generic use of the present tense (see e.g., Carlson (1982)) that is similar to that in scientific statements that express timeless truths (e.g., the fact that whales are, were and will be mammals is most naturally expressed as "Whales *are* mammals"). We thus get a preference for present tense in parafictional statements with **In**.[15] The same reasoning applies to contensive statements with **In** that report on fictions about past or future events (e.g., it is true at some point on the fictional timeline of *Star Wars* that Luke destroys the Death Star) and hence these will also display a preference for present tense.[16]

Second, the proposed analysis of **Acc** accounts for tense use in contensive statements with this operator. Contensive statements with **Acc** are analysed as indirect speech reports (i.e., reports on what a medium 'asserts'). Hence tense use in such contensive statements mirrors that of indirect speech reports. If an indirect speech report reports on a 'current' speech event (i.e., the speech verb is in present tense), then the tense use in the embedded clause mirrors that of the reported speech act. The reported speaker's tense use in turn depends on whether the time of the events described coincides, precedes or succeeds the utterance time of her statement, i.e., whether she is telling us what things *are*, *were* or *will be* like. Hence, tense use in indirect speech reports on current speech events shifts depending on whether the time of the described events coincides, precedes or succeeds the utterance time of the contensive statement. For instance, if Adeela says "Sara will be nervous" at t_1, a speech report at t_1 will mirror her tense use:

(150) Adeela asserts that Sara will be nervous.

Sentence (150) uttered at t_1 is true iff Adeela asserts at t_1 that Sara is nervous at some t such that $t > t_1$.[17]

---

[15] See Zucchi (2001) for an alternative possible world analysis of **In** that accounts for this present tense preference by switching the time of evaluation to the time of the described events.

[16] In fact, the same reasoning also applies to contensive statements with **In** that report on non-fiction (e.g., 130) which thus also display a preference for present tense (cf. Zucchi (2001)):

   (xi) In *Shadow*, Clinton only cares/? cared/# will care about sex and golf.

[17] I assume a simple analysis of "will" as a tense marker (see e.g., Prior's (1967) 'Ockhamist semantics' or Kissine (2008); Salkie (2010)). Under a modal analysis (e.g., Condoravdi (2002); Enç (1996)) "will" still has a temporal dimension and hence a modal analysis can also be incorporated into my analysis.

A contensive statements with **Acc** is analysed as a report on what a medium *asserts*. Hence it is a report on a current speech event.[18] In other words, unlike contensive statements with **In**, contensive statements with **Acc** are essentially time bound; they report on what the medium asserts *now*.[19] Likewise, tense use in contensive statements with **Acc** shifts depending on whether the events described by the medium overlap, precede or succeed the utterance time of the contensive statement. For example, since the *Star Wars* saga is about events that supposedly took place a long time ago (in a galaxy far, far away), we use past tense when we report on its content using **Acc**. For example, on either an 'extended pretence' (i.e., where we pretend that *Star Wars* is a non-fictional report) or 'export' reading (which are both hard to get here), we report that *Star Wars* asserts that Luke *destroyed* the Death Star. Hence, to the extent that (143) is acceptable, it displays a preference for past tense:

(143)   According to the *Star Wars* saga, Luke destroyed the Death Star.

Sentence (143) uttered at t_1 is true iff *Star Wars* asserts at t_1 that Luke *destroyed* the Death Star at t_1 (i.e., *destroys* the Death Star at some t such that $t < t_1$). Likewise, since a medium like the news report on protests in Amsterdam in (145) reports on events that are currently taking place and the *Star Trek* series is (amongst other things) about events that supposedly will take place in the future, we report on the content of these media using present and future tense respectively.

7.4 Outlook

7.4.1 Conclusions

In this chapter I have argued that the **In** and **Acc** operators require separate semantic analyses to account for three linguistic observations. These concern preferences for using **In** for contensive statements about fiction and **Acc** for non-fiction; the

---

[18]A complication for this comparison to indirect speech reports is that whereas the speech report about Adeela's assertion mirrors her tense use, tense use in contensive statements does not necessarily mirror the tense use in the medium itself. For instance, although a science fiction novel may be written from the point of view of the year 4020 and include the past tense statement "Mars was inhabited in 3020", it currently (in 2021) asserts that Mars *will be* inhabited in 3020.

[19]Maybe some assertions made by media can (like a person's assertions) also be past events, i.e., maybe s can sometimes also be reported on as having *asserted* that p. For instance, is it admissible on Wednesday to report on Monday's weather forecast about Tuesday with "According to Monday's weather forecast, Tuesday was going to be a great day for skiing" or should we say "According to Monday's weather forecast, Tuesday was a great day for skiing"? I leave this question to future research.

unacceptability of using **Acc** to report on implicit content (whereas **In** is fine for implicit and explicit content); and preferences for present tense in contensive statements with **In** and tense use in contensive statements with **Acc** depending on whether the events described by the medium overlap, precede or succeed the utterance time of the contensive statement.

I have proposed to adopt the Lewisian possible world analysis of parafictional statements for contensive statements with **In**: "**In** s, ϕ" is true iff in the worlds compatible with s, ϕ. I have proposed to analyse contensive statements with **Acc** as indirect speech reports: "**Acc** s, ϕ" is true iff s asserts/asserted that ϕ. Lastly, I have explained how the proposed analyses account for the three described linguistic observations.

7.4.2 Future research

'According to' and fictive closure

In the rest of this dissertation I have treated parafictional statements with **In** and **Acc** on a par. An obvious direction for further research is to explore whether and how we could incorporate the semantic analyses of **In** and **Acc** that have been proposed in the current chapter into the workspace account. This subsection gives a brief sketch of a proposal.

First of all, to model how information of the form '**Acc** s, ϕ' enters the common ground we would have to incorporate a mechanism that tags explicit content (i.e., all newly incoming information or what follows directly from that) as 'said by the source' in the workspace. In other words, apart from a workspace update with the asserted content p, we also represent the non-essential update 's said p' (see Stalnaker (2002)) at every utterance. Given our analysis of **Acc** as an indirect speech report, this amounts to an update with information of the form '**Acc** s, ϕ'. Assertive and fictive closure will thus result in the following updates: After reading a newspaper article it will be common ground that 'It is raining' and that 'According to the newspaper article, it is raining'. After reading the Harry Potter books, it will be common ground that 'In the Harry Potter books, there are wizards in England' and that 'In the Harry Potter books, according to the Harry Potter books, there are wizards in England'. This fits the proposed analysis according to which most contensive statements about fiction that are formulated with **Acc** are strictly speaking false (see section 7.3.1). As has been discussed above, there do seem to be some contensive statements about fiction formulated with **Acc** that we deem true, i.e., those that report on content viable for export. In section 4.6.2 I suggested that export is based on analogical reasoning with parafictional information. For instance, from 'In the Harry Potter books, love conquers all' we may derive that 'Love conquers all'. Here I suggest that such inferences, paired with the information

that 'Love conquers all' follows from what was explicitly stated in the Harry Potter books, would license an additional inference to 'According to the Harry Potter books, love conquers all'.

Secondly, one place where the proposed semantic analyses of **In** and **Acc** may complicate the discussion is in the analysis of bald-faced lies (see chapter 5). I leave detailed exploration of this to further research but will sketch some of the issues that may arise. In chapter 5, I argued that bald-faced lies are (as a speech act) more akin to fictional statements than to lies. However, contensive statements that report on bald-faced lies in institutionalized, rule-bound settings are most naturally formulated with **Acc** rather than with **In**:

(151)   a.   ? In the courtroom proceedings/records, Pentangeli knew no Godfather
        b.   According to the courtroom proceedings/records, Pentangeli knew no Godfather

The worry is that such reports thus mirror the preference for **Acc** that reports on non-fictional media display. However, the inappropriateness of **In** does not generalize to all bald-faced lie scenarios. For instance, we can report on the bald-faced lie in Meibauer's (2014) cheating husband scenario with **In**:

(152)   In this game that they have been playing for a few years now, he is faithful to her.

Moreover, as has been discussed, the preference for **In** in reports on fiction is not absolute, i.e., content licensed for export is fit to be reported on with **Acc**. Given that there is a truth-tracking requirement on court conversations (see section 5.4.5), the embedded content in (151b) seems apt for export.

A second, related, worry might be that updates with information of the form '**Acc** s, ϕ' take place for fiction and non-fiction alike. I argued that bald-faced lies are like fictional statements in part because they seem to involve a parafictional update as part of their essential updates; a successful bald-faced lie has to go 'on the record' and hence, like a successful fictional statement, has to make a modalized statement of the form '**In**/**Acc** s, ϕ' common ground. However, as noted above, for bald-faced lies in institutional settings (where the modalized update is really the point of the bald-faced lie), these updates are most naturally formulated with **Acc**, rather than with **In**. In light of the semantic analyses proposed in the current chapter, we may take this to suggest that 'going on the record' is something different from the Lewisian parafictional update involved in fictive closure. Rather, what is required is the (non-essential) update with information of the form '**Acc** s, ϕ' which may take place in the case of fictional and non-fictional discourse. In other words, whether you make an assertion or a fictional statement, you are always successful at 'going on the record' in this sense. However, this still would not prove that bald-faced lies in institutional settings could just as well be modelled as lies. This is because the definition of assertion still incorporates an update of the common ground with the

expressed content as essential update, i.e., a successful assertion of *p* updates the common ground with *p*. The definition of fictional statements on the other hand, does not incorporate such an update (it merely requires that the expressed content does not 'evaporate', i.e., a common ground update with '**In** *s*, *p*'). Similarly, the success of a bald-faced lie does not depend on whether the common ground is updated with the content that it expresses, e.g., Pentangeli's bald-faced lie was successful even though he did *not* update the stable common ground with the information that he did not know the Godfather (cf. section 5.4.1).

Pilot experiment on tense use

Another obvious direction of future research is to subject the proposed semantic analyses of **In** and **Acc** to empirical scrutiny. More specifically, the general picture of the use of **In** and **Acc** in contensive statements that has been presented in this chapter naturally invites experimental conformation. It would be interesting to see whether the three clusters of linguistic observations concerning preferences for fiction/non-fiction, implicit/explicit content and tense use in contensive statements with **In** and **Acc** can be experimentally verified.

To this end I have collaborated with Bimpikou on a pilot experiment that investigated tense use in parafictional statements with **In** (see Semeijn and Bimpikou (2019)). We empirically tested three variants of the Lewisian analysis of **In** that (unlike the original Lewisian analysis) specify a time of evaluation and thus make predictions about tense use. The time of evaluation in the three different analyses were respectively [1] the time of the described events (cf. Zucchi (2001)), [2] the time of fictional narration and [3] the utterance time of the contensive statement (cf. this chapter's analysis of **Acc**). To tease apart the three analyses we distinguished between individual-level predicates (e.g., 'be a detective', 'be a hobbit') and stage-level predicates (e.g., 'light a pipe', 'climb Mount Doom'). Second, we distinguished 'homodiegetic' narratives about past events in which the time of narration roughly coincides with the time of the events described in the novel (e.g., the Holmes stories, which supposedly are narrated by Holmes' friend Watson) from 'heterodiegetic' narratives about past events in which the time of narration is obviously long after the time of the described events (e.g., *The Lord of the Rings*, which supposedly are records of events in a distant past). Intuitively, at Watson's time of narration, Holmes *is* a detective and *lit* his pipe but at the time of narration of the narrator of *The Lord of the Rings*, Frodo *was* a hobbit and *climbed* Mount Doom. By contrast, at the time of the described events, Holmes *is* a detective and *lights* his pipe and Frodo *is* a hobbit and *climbs* Mount Doom. Lastly, at the utterance time of the contensive statement – assuming a uniform cross-world timeline – Holmes *was* a detective and *lit* his pipe and Frodo *was* a hobbit and *climbed* Mount Doom.

We conducted a questionnaire on Amazon Mechanical Turk. We used a 2×2 experimental design with 'type of narrative' (heterodiegetic/homodiegetic) and 'type of predicate' (individual-level/stage-level) as factors. 32 participants were shown 16 short stories (4 for every condition) each followed by 2 pairs of parafictional sentences with **In** (2 targets, 2 controls) in simple present and simple past versions. Below is an example item consisting of a homodiegetic narrative followed by target sentences with individual-level predicates:

> *Subha Datta*
> My dear friend Subha Datta is a woodcutter and every day he goes to the forest near his home to get supplies of wood. Sometimes he takes his three boys with him, and now and then, as a special treat, his two little girls are allowed to trot along beside him. Subha has told his sons that as soon as they are old enough he will give each of them a little axe of his own. The girls, he has said, must be content with breaking off small twigs from the branches he cuts down.
>
> One day, Subha told his children that none of them could come with him, for he meant to go a very long way into the forest, to see if he could find better wood there than nearer home. Vainly the boys entreated him to take them with him. "Not today," he said, "you would be too tired to go all the way. You must help your mother today and play with your sisters."

- In the fairy tale, Subha Datta encourages his daughters to chop wood.
- In the fairy tale, Subha Datta encouraged his daughters to chop wood.
- In the fairy tale, Subha Datta made a living by cutting wood.
- In the fairy tale, Subha Datta makes a living by cutting wood.

The questionnaire consisted in a four-alternative forced-choice task. Participants were given the instruction to "choose the sentence that best describes (part of) the content of the passage". Participants responses are summarized in table 1:

Table 1.

176
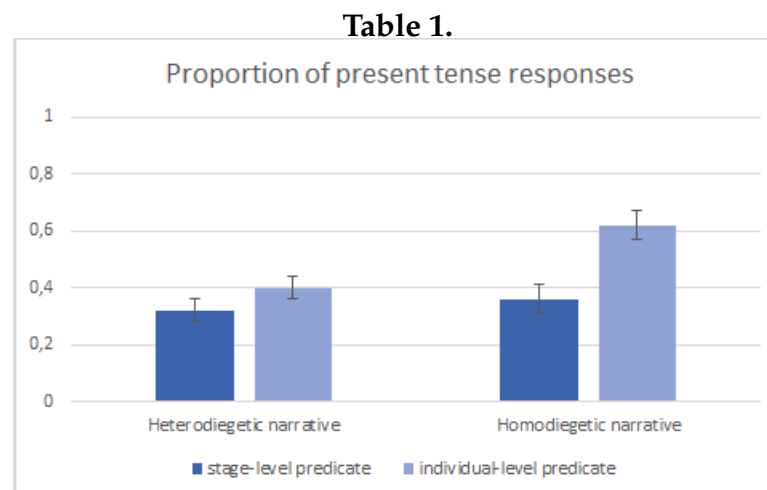
Overall results show that participants chose mainly past tense, except in the homodiegetic/individual-level condition, where they preferred present tense. This tentatively supports analysis [2] (i.e., the time of evaluation is the time of the fictional narration) and speaks against Zucchi's (2001) observation of a general preference for present tense in parafictional statements with **In**. However, since the participants answered questions about stories that were on a screen in front of them, they may have been inclined to simply copy the (narrator's) tense use rather than report their own preferences. In future experiments it would be interesting to check tense use preferences in parafictional statements about well-known narratives without showing them to the participants.

Other languages

Another potentially fruitful direction for future research is to explore to what extent the presented observations generalize to other languages. Although I focus on English language use in this dissertation, on the face of it, the **In** and **Acc** contrast also exist in other languages such as Dutch, Spanish and French:

(153) Dutch
    In De Hobbit *reist   Bilbo naar de Eenzame Berg.*
    In *The Hobbit* travels Bilbo to   the Lonely   Mountain.
    'In *The Hobbit*, Bilbo travels to the Lonely Mountain.'

(154) Dutch
    ? *Volgens   De Hobbit *reist   Bilbo naar de Eenzame Berg.*
      According-to *The Hobbit*, travels Bilbo to   the Lonely   Mountain.
    'According to *The Hobbit*, Bilbo travels to the Lonely Mountain.'

(155) Spanish
    En El Hobbit, *Bilbo viaja   a la Montaña Solitaria.*
    In *The Hobbit*, Bilbo travels to the Mountain Lonely.
    'In *The Hobbit*, Bilbo travels to the Lonely Mountain.'

(156) Spanish
    ? *Según   El Hobbit, *Bilbo viaja   a la Montaña Solitaria.*
      According-to *The Hobbit*, Bilbo travels to the Mountain Lonely.
    'According to *The Hobbit*, Bilbo travels to the Lonely Mountain.'

(157) French
    Dans Le Hobbit, *Bilbo voyage vers la   Montagne Solitaire.*
    In   *The Hobbit*, Bilbo travels to   the Mountain Lonely.
    'In *The Hobbit*, Bilbo travels to the Lonely Mountain.'

(158) French

> ? *Selon*   Le Hobbit, *Bilbo voyage vers la   Montagne Solitaire.*
>   According-to *The Hobbit*, Bilbo travels to   the Mountain Lonely.
>
> 'According to *The Hobbit*, Bilbo travels to the Lonely Mountain.'

Dutch 'in' and 'volgens', Spanish 'en' and 'según' and French 'dans' and 'selon' seem to display similar behaviour to **In** and **Acc** in English (e.g., preferences for 'in', 'en' and 'dans' over respectively 'volgens', 'según' and 'selon' for contensive statements about fiction). This suggests that there is indeed an important contrast here that is not just a peculiarity of English grammar. It would be interesting to see whether the **In**/**Acc** contrast generalizes to languages that are not Indo-European.

8 Unreliable narration and imaginative resistance

This chapter is a slightly revised version of the paper 'Extracting fictional truth from unreliable sources' which is co-authored with Emar Maier and fortcoming in OUP's *The Language of Fiction*. The most substantial difference between this chapter and the volume contribution are: first, the addition of a brief discussion of the need of some kind of revision operator for workspace updates and how this relates to the overall project of modelling fictional and non-fictional discourse uniformly in a Stalnakerian framework (section 8.2.1 and 8.2.2). Second, the removal of the introduction of the workspace account.

8.1 Introduction

It is a fiction author's prerogative to decide what's true in the fictional worlds she creates. After all, it's her words that create this world, by saying what it's like in there. When Tolkien wrote that "Frodo had a very trying time that afternoon" it automatically became true in the *Lord of the Rings* saga that Frodo had a very trying time on a particular afternoon. This line of thinking can be summed up in the principle of Authorial Authority:[1]

> Authorial Authority: If *s* is part of text *T* (and not a quotation), then the proposition expressed by *s* is true in the world of *T*.

This principle seems to hold for all fiction, and only for fiction. If a historian or journalist writes that Napoleon was 1.47m tall, this does not thereby become 'true in the world of the historical text'.[2] What the historian writes is true or false depending

---

[1]This formulation has been adapted from Badura and Berto (2019).

[2]Though see for instance Zucchi (forthcoming) for an opposing view on which any discourse or text T makes it true that 'in/according to T, ϕ' (for any ϕ in T). As has been

on what the actual world is like. At first sight then, Authorial Authority promises to help pin down what fiction is, and how it differs from non-fiction.

Unfortunately, truth in fiction is not always so straightforward. First, as has been discussed in chapter 4, fictional truth can be implicit. Lewis and many subsequent authors in philosophy and narratology have observed that there are many propositions that are true in a given fiction beyond the ones that make up the text. In the Harry Potter books, milk comes from cows, water is H_2O, and people are annoyed if you cut in line. More interestingly, the opposite is also true. There are cases where a text says that p, but that fails to be true in the world of the fiction. These cases of Authorial Authority breakdowns are likewise much discussed in both literary studies and philosophy, but in rather different terms. The first group talks about unreliable narrators, i.e., narrators that misinform or misjudge because they are trying to deceive, are prejudiced, naïve, or confused. For instance, in *The Adventures of Huckleberry Finn* the narrator, Huck, gives the following report on a dinner with the widow Douglas:

(159)   The widow rung a bell for supper, and you had to come to time. When you got to the table you couldn't go right to eating, but you had to wait for the widow to tuck down her head and grumble a little over the victuals, though there warn't really anything the matter with them.

Huck reports that Douglas grumbles over her food before eating as if she were unhappy with it but the reader realizes that Huck fails to understand that actually she was praying. Hence even though the text states that Douglas grumbled over her food, this is not true in *The Adventures of Huckleberry Finn*. Rather, it is true in the fiction that the widow prayed before taking her meal and that Huck mistook this for dissatisfied grumbling.

Somewhat independently from literary scholars' debates about unreliable narration, there is a now long-standing debate in philosophy about so-called 'imaginative resistance', a phenomenon whereby readers of a fictional text resist imagining and/or accepting a part of a story. Consider the story *Fish Tank*:

(160)   Sara never liked animals. One day, her father caught her kicking the neighbor's dog. He got really angry and she was grounded for a week. To get back at her father she poured bleach in the big fish tank, killing all the beautiful fish that he loved so much. Good thing that she did, because he was really annoying.

Readers can go along imagining a sadistic protagonist kicking her neighbor's dog and killing her father's fish, but when they arrive at the evaluative statement "Good thing that she did" they resist. Even though the text explicitly states that it was

discussed in chapter 7, I assume that there is a crucial semantic difference between 'in' and 'according to'.

good that she did this, readers report that they can't or won't imagine that this is so, nor do they accept that it is true in the story.

In sum, cases of imaginative resistance and unreliable narration alike constitute clear prima facie counterexamples to the intuitive Authorial Authority principle for fiction. But if we can't trust the author's words to give us the fictional truths, how *do* we know what's true in the story? How do readers of stories like the above figure out what the fictional world is like?

This chapter starts with a brief discussion of the need of some kind of revision operator for workspace updates and how this relates to the overall project of modelling fictional and non-fictional discourse uniformly in a Stalnakerian framework (section 8.2). We then present an extension of the workspace account that takes into account the role of the (unreliable) narrator and the phenomenon of imaginative resistance by incorporating insights from belief revision logic (section 8.3). We then apply this framework to the two concrete examples of unreliable narration and imaginative resistance above (section 8.4).

8.2 Workspace revision

8.2.1 Basic motivation

Separate from issues concerning unreliable narrators and imaginative resistance, there is in fact already a basic need for some kind of revision mechanism in the workspace account (see sections 3.3.1 and 4.4). This is because fictional stories often introduce content that is inconsistent with our shared background information. Let's consider another example of a fictional discourse to illustrate this. Suppose a reader picks up Heinlein's classic 1961 sci-fi novel *Stranger in a Strange Land*. At this point it is for instance common ground between her and Heinlein that someone named Heinlein wrote a book called *Stranger in a Strange Land*:

(161)
$$\left\langle \quad \begin{array}{|c|}\hline x, y \\\hline \text{Heinlein}(x) \quad \text{book}(y) \\ \text{wrote}(x,y) \quad \text{stranger.in.strange.land}(y) \ \dots \\\hline \end{array}^{\text{C}} \quad ,\varnothing \quad \right\rangle$$

At the start of the fictional discourse, a new workspace is opened up that is a copy of the current common ground (assuming that no relevant genre conventions are common ground). Now the workspace is updated with the information expressed by the first sentence (162) of the novel:

(162)   Once upon a time when the world was young there was a Martian named Smith.

(163) $\left\langle \begin{array}{|c} \hline \\ \hline \cdots \\ \hline \end{array} C \; , \; \begin{array}{|c|} \hline x', y', z' \quad\quad W \\ \hline \text{Heinlein}(x') \;\; \text{book}(y') \\ \text{wrote}(x',y') \;\; \text{stranger.in.strange.land}(y') \;\; \cdots \\ \text{martian}(z') \;\; \text{Smith}(z') \\ \hline \end{array} \right\rangle$
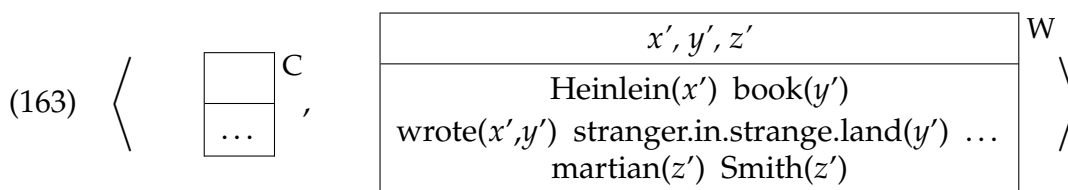
As has been discussed in chapter 4, after the fictive discourse (i.e., at fictive closure) the content expressed by (163) is quarantined from the content in the persistent common ground because it is placed under a fiction operator; after the fictional discourse it isn't common ground that there exists a Martian named Smith but it *is* common ground that 'In *Stranger in a Strange Land*, there exists a Martian named Smith'. This neatly avoids a clash in the common ground between the incoming fictional information and our background information which will definitely include some uncontroversial information about the history of space travel, hidden somewhere in the '. . . ', that conflicts with the existence of Martians named Smith.

However, remember that a new workspace is a complete copy of the entire current common ground. Hence the inconsistency remains present in the workspace during the fictional discourse and, afterwards, embedded under the fiction operator in the new persistent common ground.[3] This is unsatisfactory. We do not accept inconsistencies as true in the fiction just because incoming fictional updates clash with our background information. Rather, we need an analysis of how we easily give up background assumptions about actual space travel and the non-existence of Martians from the workspace, when confronted with a fictional text that states or entails otherwise.

The workspace account has been formulated with workspace revision (i.e., $W * p$) as an option for workspace updates.[4] This allows us to model our interpretation of fictional truth in a way that fits the Lewisian analysis of the fiction operator (i.e., everything that is common ground simpliciter is also true in the fiction unless it is contradicted by the fictional discourse). An interesting feature of this way of modelling fictional discourse is that it predicts that, since the proposition that Martians do not exist ($\neg p$) is part of the workspace at the start of the discourse, Heinlein himself also temporarily accepts $\neg p$ at the start of the discourse. He subsequently updates the workspace with p by introducing a Martian in the fictional discourse and hence $\neg p$ is removed from the workspace, i.e., Heinlein and his readers no longer accept that $\neg p$. In other words, Heinlein can fictionally state p while accepting $\neg p$. In non-fictional conversations, however, such common ground revision that is prompted by a speech act by one of the conversational participants

---

[3]In a similar way fictive opening with genre conventions (e.g., that in fairytales, dragons breathe fire) can lead to inconsistencies in the workspace.

[4]Similarly, in the unofficial common ground accounts, updates of unofficial common grounds can involve revision if we assume that a new unupdated unofficial common ground (C_{BASE}) is a copy of the current official common ground (see section 3.3.1).

seems odd, if not impossible.[5] If some speaker *a* accepts ¬*p* for the purpose of the conversation, why would they then say (something that implies) *p*? Doesn't the fact that *a* asserted *p* mean that *a* apparently did not accept ¬*p* in the first place and hence that ¬*p* wasn't actually common ground (even though the hearer might have thought it was)? Up until now, I have tacitly assumed that this is indeed how workspace updates work in fictional and non-fictional discourse. In fictional discourse what is part of the workspace does not necessarily dictate what speech acts can be made. A speaker that produces fictional discourse can accept ¬*p* but can nonetheless decide to make *p* part of their fictional discourse (e.g., decide that Martians or hobbits exist in the fiction) and hence come to, during the discourse, revise their acceptance of ¬*p*. A speaker in non-fictional discourse, however, does not seem to be free to revise what they accept in this way.

The version of the workspace account that we've been working with in previous chapters thus allows for some asymmetry between fictional and non-fictional discourse; although both assertions and fictional statements update a temporary workspace (which is a copy of the current common ground), only in the case of fictional discourse can these updates also lead to workspace revision. This shows us that it is challenging to combine total uniformity of fictional and non-fictional discourse (i.e., the desideratum we took from Matravers' theory) with complete uniformity of how speaker and hearer engage with a discourse (i.e., how discourse is modelled in Stalnaker's common ground framework that abstracts away from individual mental states). On the current analysis, a speaker's engagement with fictional discourse differs from her engagement with non-fictional discourse; in the first case she can contradict what she accepts, in the latter she cannot. This asymmetry is not contra Matravers' theory per se because it entails a difference in the *production* of fictional versus non-fictional discourse, not in the *interpretation* of the discourse. Matravers' theory really only predicts uniformity of the reader's interpretation of fictional and non-fictional discourse. In line with Matravers, the hearer's conception of the workspace can involve revision in both the case of non-fictional and fictional discourse. This is what we model in this chapter.

8.2.2 A psychologistic turn

Alternatively, we may place the asymmetry somewhere else if we analyse fictional discourse that contradicts shared background information as giving rise to a defective context, i.e., conceptions of what is commonly accepted diverge (at least momentarily). Although the reader may have accepted that Martians do not exist at the start of the discourse (and considered this to be common ground), Heinlein did

---

[5]By contrast, common ground revision that is prompted by for instance the shared perception of some event that contradicts previous shared background seems possible.

not (e.g., because he knew what kind of story he wanted to write) even though this information *is* persistent common ground. On such an analysis, workspace updates never involve revision, also not in the case of fictional discourse. The asymmetry between fictional and non-fictional discourse is now placed at the starting point of the discourse, i.e., what is commonly accepted at the start of fictional discourse is not necessarily a duplicate of the persistent common ground, although this is still the case for non-fictional discourse.

In formulating the workspace account I have up till now used DRS's as representations of Stalnakerian common grounds, i.e., speech acts have been defined in terms of how they update the common ground.[6] However, if we maintain that there is a defective context in the Heinlein case, then the workspace account (with its copying mechanism for new workspaces and revision operator)[7] only correctly represents the hearer's interpretative processes. Only the reader's conception of the workspace included the information that Martians do not exist and was subsequently revised, the actual workspace never contained this information. We could then opt for a moderate psychologistic move (see also the discussion in section 5.2) and define speech acts in terms of the hearer's conception of the workspace and common ground. DRS's would then represent represent part of the mental state of the interpreter of the discourse, viz. the interpreter's beliefs about what is common ground between herself and the speaker (i.e., what Stalnaker would call the hearer's presuppositions). In short, an assertion that *p* triggers the hearer to update her conception of what is temporarily common ground with *p* and (at the end of the non-fictional discourse) perform assertive closure, i.e., update her conception of what is persistent common ground with *p*. A fictional statement *p* (of a story *s*) triggers the hearer to update her conception of what is temporarily common ground with *p* and perform fictive closure, i.e., update her conception of what is persistent common ground with 'In *s*, *p*'.

In this chapter we remain agnostic with respect to whether common ground revision is possible in the case of fictional discourse but adopt a mentalistic representation of DRS's anyway. This is because our current aim is to model how revision

---

[6]With the notable exception of chapter 5 where I suggested that one way to model lies in a Stalnakerian framework is to define speech acts in terms of how they update the hearer's conception of the (persistent) common ground.

[7]In fact, we are not necessarily forced to make this psychologistic turn. Arguably, we could formulate a version of the workspace account purely in terms of common ground updates where it is not part of the workspace (and nobody mistakenly considers it to be part of the workspace) that Martians don't exist at the start of the Heinlein discourse. Such an account could involve starting out with a tabula rasa workspace (i.e., nothing is commonly accepted at the start of any new discourse) and accommodation of background information from the persistent common ground where necessary (see also section 3.3.1 for a brief discussion of such an approach).

processes involved in a hearer's interpretation of a discourse are uniform across fictional and non-fictional discourse. If we want to model this we have to assume that the DRS's below represent the hearer's conception of the workspace because (on either analysis of fictional discourse) in the case of non-fictional discourse, revision only takes place at this level, i.e., the actual temporary acceptance-based common ground is not revised. Consider for instance what happens when a hearer believes it to be common ground that only birds fly and then a trusted source says "You do know that bats are not birds, right?". In that case the hearer's presuppositions do not coincide with the actual workspace – the speaker may have never believed or even accepted that only birds fly – and what we're interested in capturing is how the hearer's presuppositions rather than the actual common ground are revised in light of the new information.

As for notation, we officially represent interpretation contexts as a pair consisting of (the hearer's conception of) a persistent or stable common ground and an active workspace or the empty set. As before, in workspaces we'll use primed copies of the original discourse referents (x', y',…) and in some examples we'll just display the current workspace or the stable common ground. For reasons of space we're not explicitly implementing any theory of anchoring or direct reference in DRT here (see the discussion in section 4.6.1 and e.g., Kamp (1990); Kamp (forthcoming); Hunter (2013); Maier (2009b)). We will assume that the notational convention of 'priming' all discourse referents in creating a workspace reflects a formal linking, so that if x is somehow anchored (i.e., directly referential) to Trump, so is x'.

8.2.3 Writers in fiction

Before we turn to revision, there is another striking feature of our output DRS that requires comment. According to the DRS that represents the update caused by the Heinlein discourse (163), it is part of the workspace that there is a writer named Heinlein who wrote a book named *Stranger in a Strange Land*. Hence, after fictive closure, it will become common ground that in the fictional world of the book there is a writer named Heinlein who wrote a book named *Stranger in a Strange Land*. In other words, somewhere in the fictional universe there exists not only a Martian named Smith but also some guy named Heinlein writing about Martians. Revision might help us avoid this counterintuitive consequence in cases where the content of the story conflicts with the existence of a human fiction author writing a fictional book. However, in many cases there may be nothing in the story to contradict the existence of an author with a certain name, somewhere in the background in a remote (in space and time) corner of the universe, away from the main events of the story. In the case at hand, some stable common ground assumptions about Heinlein

will perhaps have to be given up to maintain consistency.[8] For instance, it is highly unlikely that someone published a famous science fiction story about a Martian named Valentine Michael Smith, which much later, years in the future (after World War III) turned out to happen exactly as described. On the other hand, the mere fact that there was a sci-fi author (in the distant past) named Heinlein seems less controversial and may well survive revision, yielding something like the output in (163), with a discourse referent for Heinlein in the representation of the fiction. Note that this fictional Heinlein counterpart in (163) is not to be equated with the narrator of the story. For one, the narrator is temporally located at some unspecified time after the events, while this fictional Heinlein lives in the 1960's. Moreover, the fictional narrator by definition tells the story 'as known fact' (Lewis (1978)), while the fictional Heinlein, like the real one, wrote the story as pure fiction. We return to the status of the narrator in section 8.4.

Summing up, on our account some common ground facts surrounding the real-world author and book may be imported into the representation of the fictional world. Revision, especially if based on plausibility, may remove some unwanted imports (like the fact that the content of Heinlein's 1961 fiction happens to match post WW3 reality). For the remaining imports, like those depicted in (163), we'll follow Walton's (1990) lead: without any textual or other evidence to the contrary, we assume that there was a 20th century author named Heinlein in the world of *Stranger in a Strange Land*. But as this information is completely irrelevant to the events in the story, none of the fictional characters nor the narrator will ever refer to the Heinlein discourse referent, so it will quickly fade into the background, as will presumably be captured by a more realistic processing model of regular common ground updating that tracks the salience of discourse referents and/or associated conditions.

8.3 Fiction updates and belief revision

8.3.1 Introducing belief revision

In the 1980's, around the same time as linguists started developing dynamic semantics, researchers in computer science, AI, and philosophy of science started developing logical tools to describe how a system of beliefs reacts to an influx of new, possibly conflicting information. Unlike regular dynamic semantics, belief revision describes also nonmonotonic updates, i.e., removing previously established

---

[8]Consistency need not be understood as mere logical consistency. It's best thought of in terms of a gradable, context-dependent notion of possibility, coherence and/or plausibility. We won't formalise this notion here.

information from the context or belief state because the new information conflicts with those previously held beliefs.

As in dynamic semantics, there are 'representational' (or 'syntactic') versions of the theory, where a belief is a set of sentences in some logical language, and more semantic versions, where a belief is modeled as a set of possible worlds (Grove (1988)). Since we're already using the representational framework of DRT, we'll adopt a version of the former, classic belief revision theory, known as the AGM model (Alchourron et al. (1985)). More specifically, we'll adopt a version with beliefs modeled as belief bases, which contain only the agent's core beliefs, rather than the logically closed belief sets that contain everything that the agent is arguably committed to on the basis of their belief base and general principles of rationality (see Hansson (1994, 1998); Nebel (1998)). For instance, if I believe that it's raining, I'm committed to believing that it's raining or sunny, but that particular disjunction is not usually part of my core belief base.

The basic operation in AGM is contraction of a belief base K with a statement ϕ, that is, reducing the set K in such a way that it no longer entails ϕ. AGM spells out a number of postulates to axiomatize well-behaved contraction operations. One way of constructing such a well-behaved contraction operation is on the basis of a given 'epistemic entrenchment' order: $\phi < \psi$ iff ϕ is less entrenched than ψ, i.e., ψ has more epistemic worth (e.g., because it derives directly from a trusted knowledge source) and therefore is less easily given up than ϕ. Natural and moral laws for instance may be considered to be more entrenched than concrete contingent facts, especially if based on hearsay rather than direct perception. An agent's belief base is fully characterized by a set of statements K and an entrenchment ordering $<$ (again, satisfying certain axioms of rationality, like transitivity and the fact that logical consequences of ϕ are at least as entrenched as ϕ itself, Gärdenfors (1988)) on the set of well-formed formulas of the language. Contracting K with (a non-tautology) ϕ (notation: $K \div \phi$) now means that we chose a $K' \subseteq K$ such that K' does not entail ϕ. Epistemic entrenchment helps us single out an optimal such K', for instance with the following definition of entrenchment-based contraction:

(164) $K \div \phi = \{\psi \in K \mid \phi < (\phi \vee \psi), \text{ or } \phi \text{ is a tautology}\}$

Gärdenfors and Makinson (1988) show that the above definition generates a well-behaved contraction operation, obeying all their rationality postulates. However, when we consider only finite belief bases, rather than logically closed belief sets, assuming a full entrenchment order on the entire language seems like overkill. Hence, Williams (1994), for instance, introduces the notion of an 'ensconcement', which is essentially a finite entrenchment on the formulas in the base. We refer to Nebel (1998) for in-depth study of entrenchment and related notions (e.g., 'prioritized base revision') applied to belief bases rather than belief sets. Below we'll continue to use the familiar term 'epistemic entrenchment', requiring only an intuitive understanding of how an entrenchment relation on a finite set of statements K can guide

the process of contracting K with ϕ, by letting it eliminate from K as few as possible of the least entrenched conditions as needed to avoid entailing ϕ. Concretely, we just start from the lowest rank and then move up to the next if that doesn't help us get rid of ϕ.

Once we have contraction, AGM defines belief revision with p as the process of first contracting with $\neg p$ and then adding ('expanding with') p. But since we're dealing with belief bases, which, unlike belief sets, need not be consistent, we can also do it the other way around: first expand with p and then contract with $\neg p$. Either way, the resulting belief base always entails p, i.e., new incoming information trumps all previous beliefs. In so-called 'semi-revision' we level the playing field and treat old and new information on a par, with only epistemic entrenchment as the guiding factor. Formally, this amounts to adding p and then removing the contradiction:
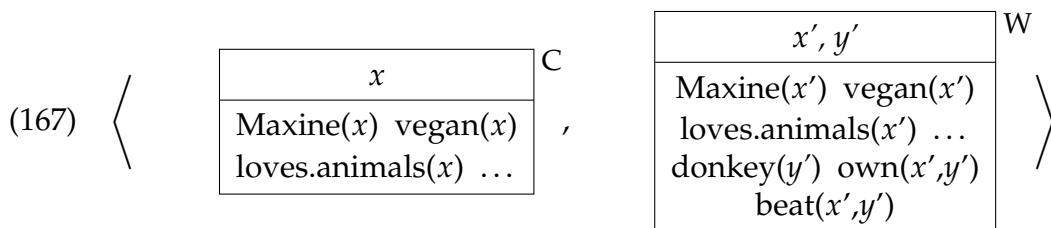
(165) $(K \cup \{p\}) \div \perp$

Below we implement this kind of revision in our DRT update mechanism so we can deal with the pervasive nonmonotonic updating required to incorporate fictional statements that contradict the initial common ground copy. See Badura and Berto (2019) for a similar application of belief revision to fiction, but in a more semantic possible (and impossible) worlds approach.

8.3.2 Belief revision in the DRT workspace

First, consider a mini-discourse that leads to an inconsistent workspace in the domain of non-fictional conversation. Consider a conversation between a speaker and a hearer who believes it to be common ground that there is a person called Maxine who is vegan and who loves animals. We open a new workspace with an exact copy of this information. The speaker now says (166), resulting in an updated workspace (167).

(166) Maxine owns a donkey. She beats it.

(167)
$$
\left\langle
\begin{array}{|c|}
\hline
x \\
\hline
\text{Maxine}(x) \ \ \text{vegan}(x) \\
\text{loves.animals}(x) \ \ \ldots \\
\hline
\end{array} \ ^{\text{C}}
\ , \
\begin{array}{|c|}
\hline
x', y' \\
\hline
\text{Maxine}(x') \ \ \text{vegan}(x') \\
\text{loves.animals}(x') \ \ \ldots \\
\text{donkey}(y') \ \ \text{own}(x',y') \\
\text{beat}(x',y') \\
\hline
\end{array} \ ^{\text{W}}
\right\rangle
$$

Given certain background assumptions about the relation between loving and beating, and donkeys being animals, conveniently hidden in the '...' in (167), the conjunction of loving animals and beating donkeys may well entail a contradic-

tion.[9] Depending on for instance how much the hearer trusts her own background information about Maxine (and about donkey keeping) and how reliable she takes the speaker to be, she will then want to revise the workspace, giving up some piece of information in order to restore consistency. We can implement the central insights from AGM belief revision introduced above to model this.

Note first that instead of belief bases we now have DRS's (pairs of sets of discourse referents and DRS conditions). To incorporate the entrenchment order we number the DRS conditions and add a third DRS compartment specifying a partial order on the conditions via these number labels. This models the epistemic entrenchment of the various bits of information that make up the DRS and thereby guide the process of resolving inconsistencies.[10] Concretely, the stable common ground representation at the start of the vegan discourse may now look like this, modeling a situation where the hearer quite strongly supposes it to be common ground between himself and the speaker that Maxine indeed bears that name, and is less invested in it being common ground that she's vegan and loves animals:

(168)

x	C
1:Maxine(x) 2:vegan(x)	
3:loves.animals(x) ...	
$1 > \{2,3\}$	

Second, we assume that instead of a classic dynamic DRS update of the workspace with incoming utterance information we perform semi-revision; new information is added to the workspace, presuppositions are resolved, and the new information is assigned a position in the epistemic entrenchment ordering. We then contract with Falsum to remove any contradictions entailed by the updated DRS K' (as per (165)). Intuitively, we do this by eliminating as few as possible of the least entrenched conditions until the DRS is consistent again.

To continue our example, let's assume that it's common ground that the speaker is a close friend of Maxine, and appears to have no reason to deceive the hearer. We can capture this by placing the new information relatively high, say just below the information that the person under discussion bears the name Maxine, but above the animal-loving-veganism. (From here on we'll display only the current workspace, with primed discourse referents, leaving out the stable common ground DRS.)

---

[9]We might eventually want to incorporate plausibility and/or coherence metrics into our model and replace 'contradiction' with 'low plausibility/coherence', i.e., a score below a certain contextually determined plausibility threshold.

[10]For convenience, we number only the conditions, not the discourse referents. This is just a technical hack: if a DRS is inconsistent we can always restore consistency by merely removing conditions, because in the extreme case a discourse referent that does not occur in any conditions doesn't actually contribute any information (except that the domain is non-empty). The set of DRS conditions thus plays the role of the belief base.

(169)   Maxine owns a donkey.

(170)

x', y'	W
1:Maxine(x') 2:vegan(x') 3:loves.animals(x') ... 4:donkey(y') 5:own(x',y')	
1 > \{4, 5\} > \{2, 3\}	

When we encounter the next sentence we again start by updating the DRS, resolving anaphora, and extending the epistemic entrenchment ordering. Let's say the current speaker's contributions about Maxine are all assumed to be of equal epistemic value:

(171)   She beats it.

(172)

x', y '	W
1:Maxine(x') 2:vegan(x') 3:loves.animals(x') ... 4:donkey(y') 5:own(x',y') 6:beat(x',y')	
1 > \{4, 5, 6\} > \{2, 3\}	

Given some very general, uncontroversial background assumptions (hidden in the '...' or kept in a separate encyclopedic knowledge compartment of a full representation of context, Kamp (2018)) about the relationships between animal loving and donkey beating, this DRS is arguably inconsistent. And even if not logically inconsistent it's questionable as a representation of the common ground, as it's unlikely that Maxine is both an animal lover and a donkey beater. The least entrenched conditions are 2 and 3. Elimination of condition 3 ('loves.animals(x')') is already sufficient to make the DRS consistent again:

(173)

x', y'	W
1:Maxine(x) 2:vegan(x') ~~3:loves.animals(x')~~ 4:donkey(y') 5:own(x',y') 6:beat(x',y')	
1 > \{4, 5, 6\} > \{2,~~3~~\}	

After processing this mini-discourse, we perform assertive closure, which turns the workspace in (173) into the new, updated common ground.

8.3.3 Cautious update

In the above instance of semi-revision the hearer considered the speaker to be very reliable, i.e., the incoming information was assigned a place high in the epistemic entrenchment ordering. We have thus essentially modeled a non-monotonic generalization of what Eckardt (2014) calls a Trust Update, i.e., we add the information of the speaker's utterance directly to the common ground. However, as Eckardt also notes, we do not always trust the speaker. Suppose that the hearer actually knows Maxine really well and assumes it is definitely common ground that Maxine is vegan and also definitely loves animals. However, she also knows that the speaker may not be very reliable when it comes to Maxine, so what she says may be based on shaky assumptions, or lies, and hence shouldn't automatically become established common ground. In terms of entrenchment, the incoming information about Maxine's donkey, conditions 4-6 in the pre-contraction DRS (172), now instead dangle at the bottom of the entrenchment hierarchy: $\{2,3\} > \{4,5,6\}$. Since this DRS again represents an inconsistent common ground, we need to remove a low ranked condition to restore consistency. On the current ranking, revision will simply cancel the latest update, 6:

(174)

x', y'	W
1:Maxine(x') 2:vegan(x') 3:loves.animals(x') … 4:donkey(y') 5:own(x',y') ~~6:beat(x',y')~~	
$1 > \{2,3\} > \{4,5,\sout{6}\}$	

In other words, the speaker's last utterance ('she beats it') is inconsistent with previous, more entrenched information and is therefore essentially ignored by the hearer.

This not quite right. Especially when we'll be trying to extract meaning from unreliable narrators in fiction, we can't completely ignore speakers just because we don't trust them. Even if we do not trust a speaker's assertion that p, we can still extract valuable information from the utterance, viz. the information that the speaker themselves believed that p, or at the very least, in case they are lying, that they asserted that p and are thereby committed to p. Although the distinction between these two kinds of unreliability is important, not least in making sense of literary unreliable narrators, we'll lump them together here and use the uniform weak Stalnakerian attitude verb of acceptance ($\approx$ treating a proposition as true, see chapter 2) to describe the hedged information we can extract from an unreliable speaker.

We suggest incorporating this Cautious Update (Eckardt (2014)) into the non-monotonic workspace update mechanism: whenever semi-revision leads us to

cancel part of the semantic contribution of the current speech act, we instead replace the offending condition ϕ with a suitably hedged version under a modal operator: $ACCEPT_x\phi$, depending on whether the hearer considers the speaker to be misinformed or deceptive, with x a discourse referent picking out the current speaker. In this case we'll assume there's been a discourse referent s' and a condition 0 representing the speaker in the workspace (and stable common ground) all along. Note also that the hedged condition, here 7, will be assigned a new, typically higher, place in the entrenchment ranking.

(175)

x', y', s'	W

0:speaker(s') 1:Maxine(x')
2:vegan(x') 3:loves.animals(x') …
4:donkey(y') 5:own(x',y')
~~6:beat(x',y')~~
7:ACCEPT$_{s'}$ $\boxed{\text{beat}(x',y')}$
$\{0,1\} > \{2,3\} > 7 > \{4,5,\!\not\!6\}$

Interestingly, because the DRS represents the hearer's beliefs about what is common ground between her and the speaker, once the DRS is updated with the hedged version (i.e., that the speaker accepts that Maxine beats her donkey) we can't really keep the information that Maxine loves animals. Although it is not strictly contradictory or even prima facie implausible that Maxine loves animals while the speaker accepts that she beats her donkey, it cannot be common ground between speaker and hearer that this is so. To see this, note that ϕ being common ground entails that it is commonly known that everyone (so, in particular, the speaker) accepts ϕ. Thus, the assumption that (175) is common ground will entail that it is commonly known that the speaker accepts (175). Since (175) is essentially a conjunction of the various conditions therein and common ground and acceptance operators distribute over conjunction, it follows that (i) it is common ground that condition 7 holds (i.e., it is common ground that the speaker[11] accepts that Maxine beats her donkey), and (ii) it is common ground that the speaker accepts condition 3 (i.e., it is common ground that speaker accepts Maxine loves animals). This would mean that it is now common ground that the speaker is inconsistent, which is clearly not the case here (regardless of whether she's lying or confused about the facts, she probably doesn't believe in contradictions). Instead, we have to give up condition 3, the assumption that Maxine loves animals. Note that the hearer herself probably really believed condition 3 to be true, and even took it to be common ground. In

---

[11]Note that at this point we rely on the fact that s' denotes the actual speaker, i.e., our assumption that s' is linked to s which is in turn anchored to the actual current speaker. See our brief discussion of direct reference and workspace linking in 8.2.2.

fact her personal belief in condition 3 will likely remain unaffected, but after the speaker's assertion it can no longer be considered part of the common ground, for it has become clear that they don't share a commitment to this information. The end result thus will be:

(176)

$$\begin{array}{|c|}\hline x', y', s' \qquad\qquad \text{W} \\\hline \text{0:speaker}(s') \;\; \text{1:Maxine}(x') \\ \text{2:vegan}(x') \;\; \text{3:loves.animals}(x') \ldots \\ \text{4:donkey}(y') \;\; \text{5:own}(x',y') \\ \text{6:beat}(x',y') \\\hline \text{7:ACCEPT}_{s'} \;\boxed{\text{beat}(x',y')} \\\hline \{0,1\} > \{2,3\} > 7 > \{4,5,6\} \\\hline\end{array}$$

As we will see below, a similar reasoning applies to the case of cautious updating in fiction.

We can incorporate all the above reasoning into a definitive nonmonotonic, cautious workspace update algorithm along the following lines:

(177)   Update a workspace K with a preliminary DRS representation ϕ (of incoming utterance), notation: $K + \phi$

   a.   expansion: $K \uplus \phi$ = merge K with ϕ, resolve all anaphora and presuppositions, and extend the epistemic entrenchment ranking to the new conditions.

   b.   contraction: $(K \uplus \phi) \div \bot$ = remove as many low ranked conditions from $K \uplus \phi$ as needed to ensure that $CG_E(K \uplus \phi)$ is consistent (where E denotes in every world the set of people engaged in the discourse in that world, and $CG_E K$ entails $\forall x \in E(ACCEPT_x K), \forall x \in E(ACCEPT_x(\forall x \in E(ACCEPT_x K))), \ldots$)

   c.   caution: for any condition ψ added in the expansion phase but subsequently removed in the contraction phase, update with the corresponding hedged condition $ACCEPT_{s?}\psi$ (where '$s?$' is an anaphor that needs to be bound to the current speaker), i.e., $((K \uplus \phi) \div \bot) + \boxed{ACCEPT_{s?}\psi}$

Continuing with our example. After assertive closure, (176) will become the stable common ground, i.e., Maxine the vegan animal-lover owns a donkey and the speaker accepts (perhaps even believes, but in any case commits herself to it by asserting it) that Maxine beats her donkey. Note again how this correctly captures the hearer's conception of the common ground, but not the speaker's, because the speaker might actually consider it to be common ground that Maxine beats her donkey, nor does it accurately capture either the speaker's or hearer's private beliefs about Maxine. To align all these mental states and conceptions of the common

ground, the hearer would have to manifest her distrust and renegotiate an aligned common ground with the speaker.

8.4 Interpreting fiction

Now that we have introduced our basic framework we will apply it to the interpretation of fiction. We show how various interpretation strategies emerge from our workspace account, allowing us to model the various ways of constructing imaginative story worlds from fictions featuring impersonal and personal, reliable and unreliable narrators.

We start with a simple face value interpretation of a fictional text with a reliable, impersonal narrator. Then we turn to cases involving Cautious Update triggered by unreliable narrators as in *The Adventures of Huckleberry Finn*. Lastly we apply our framework to a typical case of imaginary resistance.

8.4.1 Authorial Authority revisited: face value interpretation by shielding

The algorithm we sketched in (177) for updating a workspace works the same with fiction as with non-fiction. In section 8.1 we identified one prima facie difference between fiction and non-fiction: the principle of Authorial Authority (i.e., whatever the text asserts, is true in the fiction). We can reformulate this principle now in terms of epistemic entrenchment: conditions derived from interpreting a fiction outrank pre-existing conditions in the workspace (derived from copying the stable common ground and genre conventions). In other words, when reading fiction we simply interpret the incoming information as extremely reliable.

For instance, reconsider the workspace we get by expanding an input context with the Heinlein opening passage, i.e., (163), but now with an epistemic entrenchment ordering on its conditions. For expository purposes we've also included a condition 5 to abbreviate the previously hidden cluster of commonly known scientific facts about the evolution of life in our solar system that would clash with the existence of a martian named Smith.

(178)

x', y', z'	W
1:Heinlein(x') 2:book(y')	
3:wrote(x',y') 4:stranger.in.strange.land(y')	
5:mars.is.uninhabited …	
6:martian(z') 7:Smith(z')	
$\{6,7\} > 5 > \{1,2,3,4\}$	

In the context of a factual inquiry, condition 5 would outrank almost anything you can tell me. If you tell me, factually, you saw a Martian the other day, I'd sooner assume you're joking,[12] speaking metaphorically, or lying, than remove some of the basic scientific assumptions underlying 5 from the workspace.[13] When it's understood as fiction, information deriving from the text may well outrank basic science, as illustrated in (178). Those fictional statements are effectively 'shielded' from contraction, i.e., they will never be given up, even if they are inconsistent with some other seemingly uncontroversial statement that is part of our general background knowledge.

In (178), eliminating one or several of the least entrenched conditions (1-4) will not make the workspace consistent. Next up in the epistemic entrenchment ordering is condition 5, whose elimination does make the workspace consistent. We end up with a workspace where some facts about human space travel and life in our solar system are no longer valid.[14] Unlike the destructive copy operation of Assertive Closure, Fictive Closure however doesn't remove these retracted assumptions from the stable common ground. If after reading a few more pages we close this workspace, the resulting common ground is as follows:

(179)

x, y	C
1:Heinlein(x) 2:book(y) 3:wrote(x,y) 4:stranger.in.strange.land(y) 5:mars.is.uninhabited …	

6:$\Box_y$
x', y', z'
1:Heinlein(x') 2:book(y') … ~~5:mars.is.uninhabited~~ … 6:martian(z') 7:Smith(z') …
$\{6,7\} > \!\!\!\!\! \diagup \, 5 > \{1,2,3,4\}$

$$5 > \{1,2,3,4,6\}$$

We call the interpretation strategy of assigning the highest possible epistemic rank to information deriving from a text, a face value interpretation of a fictional text

---

[12]Perhaps joking is a form of narrative fiction, in which case we'd no longer be engaging in factual inquiry but fiction.

[13]Talking to a young child, crazy person, or time traveller may make me remove 5 from the workspace, if it becomes clear to me that these basic facts are really not common ground between us. We've discussed the reasoning behind such revisions triggered by Cautious Update in the vegan example in section 8.3.3.

[14]Recall, Cautious Update is not triggered because we're retracting only old information (see (177)).

Matravers (2014); Altshuler and Maier (2020b); Badura and Berto (2019), i.e., an interpretation in line with the principle of Authorial Authority.

Face value interpretations are appropriate in many cases (e.g., it gives us the desired result that it is not true in *Stranger in a Strange Land* that Martians don't exist). However, as pointed out in the introduction, in some cases even in fiction we cannot blindly trust the speaker. In the remainder of this paper we will discuss the interpretive processes at work in making sense of such narratives, starting with unreliable first-person narration.

8.4.2 Unreliable narrators

Consider again our central example of unreliable narration from *The Adventures of Huckleberry Finn*, abbreviated from (159):

(180) The widow rung a bell for supper . . . When you got to the table you couldn't go right to eating but you had to wait for the widow to tuck down her head and grumble a little over the victuals . . .

Let's assume that before engaging with the novel the reader takes it to be common ground between her and Twain that the latter produced a novel called *The Adventures of Huckleberry Finn* and a lot of other background information, including some information that entails that when people in 19th century Missouri bow over their food and mumble a bit before eating they are saying a prayer. As before we represent this rather trivial cultural background assumption as a deeply entrenched condition in the stable common ground:

(181)

x, y	C
1:Twain(x) 2:author(x)	
3:adventures.of.huckfinn(y) 4:wrote(x,y)	
5:mumble.before.dinner.is.prayer . . .	
$5 > \{1, 2, 3, 4\}$	

When we open the book and start reading, we open a copy of (181) as our workspace. Unlike the Heinlein story, the story is written in the first person, featuring Huck Finn as the narrator. Huck is obviously a different person from (the fictional counterpart of) the author Twain (represented by the linked discourse referent x'). This means the reader quickly accommodates an extra discourse referent for the first-person pronouns, representing a narrator named Huck Finn who we take to be asserting the sentences that constitute the story, and who is thereby committed to their truth. Twain, with whom we are maintaining a common ground, is merely 'presenting' or 'reproducing' Huck's assertions and is *not* committed to their truth (nor is his fictional counterpart, for that matter). By contrast, note that in the Heinlein story

there was no 'I', no sign of any personalized character telling the story, and hence no need to accommodate an extra discourse referent for a asserter/narrator.

Since Huck is evidently a naive young boy, we don't always trust his assertions, just as we don't always trust our face-to-face interlocutors. When a conflict arises between the reader's background knowledge, as imported from the common ground, and the text, we might therefore want to revise the contribution of the text rather than the background. In other words, since the text is considered to be the assertions of a child, the semantic contributions of the text should not generally end up at the top of the epistemic entrenchment ranking. More generally, for first-person narratives, i.e., narratives where we accommodate a discourse referent for a first-person speaker, we relax the principle of Authorial Authority, by giving up the requirement that new information is shielded from revision by automatically ranking it at the top.

When the reader arrives at the mumbling passage, (180), she first updates the workspace with the unproblematic fictional statements (e.g., that Douglas rung the bell and that Huck had to wait for dinner etc) that do not conflict with any background information. When we get to the statement that Huck had to wait for Douglas to grumble over her food, a conflict arises. The general knowledge that when people bow over their food and speak before dinner they are praying implies that Douglas was mumbling a prayer rather than grumbling over the food.

(182)

x', y', u', s'	W
1:Twain(x') 2:author(x')	
3:adventures.of.huck.finn(y') 4:wrote(x',y')	
5:mumble.before.dinner.is.prayer …	
6:Huck(s') 7:narrator(s') …	
8:Douglas(u') 9:rang.bell(u') 10:wait.for.dinner(s')	
11:grumbling.over.food(u') …	
$5 > \{6,7,8,9,10,11\} > \{1,2,3,4\}$	

Eliminating one or several of the least entrenched conditions (1,2,3 or 4) will not make the workspace consistent. Hence we move up in the epistemic entrenchment ordering. Eliminating only condition 11 (Douglas grumbled over the food) will make the workspace consistent. But note that this is one of the new contributions, so we have to be cautious and update with the hedged variant, i.e., that the speaker asserted, and therefore accepted as true, that Douglas grumbled over the food.

(183)

x', y', u', s'	W
1:Twain(x') 2:author(x') … 5:mumble.before.dinner.is.prayer … 6:Huck(s') 7:narrator(s') … 8:Douglas(u') 9:rang.bell(u') 10:wait.for.dinner(s') ~~11:grumbling.over.food(u')~~	
12:ACCEPT$_{s'}$ $\boxed{\text{grumbling.over.food}(u')}$	
$5 > \{6,7,8,9,10,\cancel{11},12\} > \{1,2,3,4\}$	

Interestingly, unlike with the cautious update in the vegan case from section 8.3.3, the workspace can retain the generic background information that when people bow over their food and softly mumble before dinner they are praying, even after the cautious update with the hedged information that the speaker, Huck Finn, takes Douglas to be grumbling. What's more, since we eliminate as little as possible in making the workspace consistent again – even though we eliminate condition 11 – we maintain information such as that Douglas mumbled over her food before dinner. This, in combination with condition 5, implies that the workspace is updated with the information that Douglas was praying before dinner (something that was never mentioned in the fictional text).

These kinds of inferences are possible in the workspace because the workspace derives from the (reader's conception of the) stable common ground between Twain and his readers, not that between Huck Finn, the relevant speaker whose speech acts are being interpreted, and his (fictional) narratee. Hence, after a cautious update, the workspace that we copied off this common ground may still retain what Twain and his readers commonly accepted, if only temporarily while entertaining the content of the fiction at hand. This explains how in the case of fiction, a reader and author can have what Booth (1961) calls a 'communion behind the narrator's back'; it is as if the reader and Twain are listening to the narrator together and it is common ground between them that the narrator believes something false. In fact, such a 'communion' may also take place in non-fictional discourse. For instance, in case a 'speaker' reads out a (non-fiction) letter written by someone else, the hearer will also have to accommodate an extra discourse referent for the letter writer who, unlike the current speaker, actually asserted the content of the letter and is committed to its truth. In such a scenario there might also arise a communion behind the letter writer's back where it is common ground between speaker and hearer that the letter writer is unreliable.

If we apply fictive closure to (183) we update the stable common ground with this information embedded under the relevant fiction operator. Hence after reading this passage the reader takes it to be common ground between her and Twain and other engagers that in *The Adventures of Huckleberry Finn* the widow Douglas rung

a bell for supper and prayed before dinner, and that Huck (mistakenly) took her to be grumbling unhappily over her food.

Generalizing beyond this particular example it is worth stressing that cultural or other background information, like our condition 5 about prayer and dinner customs in (183), need not outrank incoming fictional statements, even in a first-person narrative (for instance if the narrator were judged more mature and reliable, or if we're dealing with a fantasy story about an alien civilization without religion or prayer). How highly entrenched certain background information is relative to incoming textual information heavily depends on genre, i.e., with respect to what clusters of facts the fiction is expected to be realistic (Ryan (1991)). For instance, if we are aware of the genre of *A Christmas Carol* as a 19th century gothic horror story, we may expect it to be realistic with respect to geographical facts (i.e., where countries and cities are located) but not necessarily with respect to all taxonomic facts (i.e., what species exist and how they are individuated). Therefore a statement such as "He rode into London, the capital of France" would trigger an unreliable narrator interpretation (we are reluctant to cancel our geographical background information that London is the capital of England). But a statement such as "[H]e looked the phantom through and through, and saw it standing before him" will not trigger an unreliable narrator interpretation in the context of *A Christmas Carol*. On the other hand, knowledge of for instance crime novel genre conventions may lead us to expect *A Study in Scarlet* to be realistic with respect to both geographical and taxonomic facts. Hence the same two fictional statements as part of *A Study in Scarlet*, would both trigger an unreliable narrator interpretation where Watson is hallucinating or otherwise mistaken about the location of London and the existence of ghosts.

8.4.3 Imaginative resistance

In section 8.1 we discussed a seemingly related case of Authorial Authority breaking down, viz. in the story called *Fish Tank* which is a typical example of what philosophers call Imaginative Resistance (Liao and Gendler (2016)):

(184)   Sara never liked animals. One day, her father caught her kicking the neighbor's dog. He got really angry and she was grounded for a week. To get back at her father she poured bleach in the big fish tank, killing all the beautiful fish that he loved so much. Good thing that she did, because he was really annoying.

Face value interpretation

Let's explore what a face value interpretation of this story would look like. Suppose that the reader of the story believes it is common ground between her and the

author that killing animals (for no good reason) is wrong. Moreover, she takes this moral law to be quite deeply entrenched, i.e., she'll be quite reluctant to give up the assumption that it is part of the established common ground between her and the author on the basis of new information and experiences. At the start of the discourse the workspace is a copy of this common ground:

(185)

x', y'	W
1:author(x') 2:fish.tank(y') 3: wrote(x',y')	
4:killing.animals.is.wrong . . .	
$4 > \{1, 2, 3\}$	

The workspace is updated with the statements that Sara never liked animals, kicked the dog, got grounded, and poured bleach in the fish tank. Since there doesn't seem to be a personal, first-person narrator the reader might assume an impersonal, omniscient narrator and, per Authorial Authority, assign these statements the highest possible ranking in the epistemic entrenchment ordering. Now we expand the workspace with the statement that Sara did a good thing, and still treat that as equally ranked with the rest of the text. We get a conflict between the moral law about killing animals and the fact that it's a good thing she killed her father's fish, which will be resolved by eliminating the moral law.

(186)

x', y', u', v'	W
1:author(x') 2:fish.tank(y') 3:wrote(x',y')	
~~4:killing.animals.is.wrong~~ . . .	
5:Sara(u') 6:father(v',u')	
7:¬ like.animals(u') 8:pour.bleach(u')	
9:did.good(u')	
$\{5, 6, 7, 8, 9\} > 4 > \{1, 2, 3\}$	

Fictive closure leads to an output where the reader takes it to be common ground in the community of engagement that killing animals is wrong, and there's a story called *Fish Tank* in which a girl called Sara poured bleach in a fish tank because she's annoyed and this was a good thing, since apparently in this fictional world moral laws are such that killing animals for trivial reasons is okay.

Non-face value interpretations through cautious update

Intuitively the face value interpretation is unsatisfactory for *Fish Tank*; many readers – even non-philosophers – feel that even though it is explicitly stated that Sara did a good thing, she actually did *not* do a good thing in the fictional world she inhabits. Empirical studies like Kim et al. (2018) and Altshuler and Maier (2020a) support

this intuition, suggesting that fictional statements, even in third-person omniscient narrations, are not always shielded from contraction and that moral truths are really quite hard to give up.[15] The flexibility of our model allows us to model this alternative interpretation by simply adopting a different entrenchment ranking strategy on the incoming textual information.

Concretely, for the non-face-value reader, the initial updates are the same: incoming information gets a high rank by default, as we're dealing with fiction and there is no reason to distrust the fictional speaker, in fact no reason to assume the presence of a narrating source at all. When we get to the final statement the reader may reconsider this assumption, because it will lead to the unwanted face value interpretation. So instead let's rank the final sentence contribution below the obviously deeply entrenched moral law. Our update algorithm, as spelled out in (177), then leads to the elimination of the final contribution followed by a hedged update:

(187)

x', y', u', v'	W
1:author(x') 2:fish.tank(y') 3: wrote(x',y')	
4:killing.animals.is.wrong ...	
5:Sara(u') 6:father(v',u')	
7:¬ $\boxed{\text{like.animals}(u')}$ 8:pour.bleach(u')	
9:~~did.good(u')~~ 10:ACCEPT$_{s?}$ $\boxed{\text{did.good}(u')}$	
$\{5,6,7,8,10\} > 4 >\!\!\!\!- 9 > \{1,2,3\}$	

One difference with our previous examples of cautious updating (Maxine the vegan and Huckleberry Finn) is that the indexical/anaphoric element in the hedging operator ACCEPT$_{s?}$ '*the current speaker* accepts that ϕ' has no obvious antecedent; there is no discourse referent for a salient current speaker in the workspace universe, as there has been no sign of a first-person narrator.[16] And it is not at all clear who this speaker should be, i.e., who is it that asserts and thereby commits themselves to Sara doing a good thing?

---

[15]For whatever reason, see e.g., Gendler (2000), Yablo (2002), or Weatherson (2004) for philosophical investigations of what makes moral truths especially hard to give up, and Andow (2019) for an empirical investigation.

[16]We might take the evaluative construction *good thing* itself as lexically presupposing a first-person judging agent, i.e., 'good' = 'good according to me' (Altshuler and Maier (2020a)). We refrain from going down this path to stay neutral with respect to the semantics and pragmatics of evaluative terms. On our account it is the cautious update itself that pragmatically triggers the accommodation of a fictional first person committed to this moral judgement.

A first option is to take the actual author, represented by a discourse referent x in the stable common ground and hence by the linked discourse referent x' in the workspace, to be the speaker:

(188)

$$
\begin{array}{|c|}
\hline
x',\, y',\, u',\, v' \qquad \text{W} \\
\hline
\text{1:author}(x')\ \ \text{2:fish.tank}(y')\ \ \text{3: wrote}(x',y') \\
\text{4:killing.animals.is.wrong} \dots \\
\text{5:Sara}(u')\ \ \text{6:father}(v',u')\ \dots \\
\hline
\text{10:ACCEPT}_{x'}\ \boxed{\text{did.good}(u')} \\
\hline
\{5\dots10\} > 4 > \{1,2,3\} \\
\hline
\end{array}
$$

As the resulting interpretation involves *de re* reference to the actual author, just like *1984* involves a *de re* reference to the actual London, we'll call this the *de auctore* reading. This reading is reminiscent of what Gendler (2000) calls a 'pop-out' interpretation in that the value judgement in the closing statement represents the evaluating of (a fictional counterpart of) the actual author.[17] Now, note that in this de auctore interpretation the source of the (fictional) speech acts, who is committed to the objectionable content, is the actual speaker, so at this point, before the fictive closure operation, the situation is entirely parallel to the non-fiction case involving cautious updates. In particular, the reasoning that made us retract the information that Maxine loves animals from (175) to (176) applies here, yielding (189).

(189)

$$
\begin{array}{|c|}
\hline
x',\, y',\, u',\, v' \qquad \text{W} \\
\hline
\text{1:author}(x')\ \ \text{2:fish.tank}(y')\ \ \text{3: wrote}(x',y') \\
\sout{\text{4:killing.animals.is.wrong}} \dots \\
\text{5:Sara}(u')\ \ \text{6:father}(v',u')\ \dots \\
\hline
\text{10:ACCEPT}_{x'}\ \boxed{\text{did.good}(u')} \\
\hline
\{5\dots10\} \cancel{>} 4 > \{1,2,3\} \\
\hline
\end{array}
$$

A reader may resist a de auctore interpretation for various reasons.

First, in some stories the narrator is a richly personalized character in the story (e.g., Huckleberry Finn). Taking the actual author (e.g., Mark Twain) as the story-teller in the fictional world would thus require making a host of (metaphysically drastic) revisions to our conception of who the author is in order to engage with the fiction. Although we do not want to exclude such an interpretation in principle, it

---

[17]Though we haven't spelled out the semantic details of direct reference and linking, we assume that after fictive closure the linked discourse referents x and x' either still represent the same individual occurring in different worlds (with some different non-essential properties in each world), or they represent counterparts of each other.

seems more intuitive to work with two separate conceptions of two separate people (i.e., author and narrator) instead, as we saw in our analysis of the Huckleberry Finn example in Section 8.4.2.

Second, as shown above, a de auctore interpretation of The Fishtank (189) will force us to remove 4 from the workspace. Hence, the information that killing animals is wrong will not end up embedded under the 'In *Fish Tank*' operator at fictive closure. The reader will thus not consider it to be common ground that in *Fish Tank*, killing animals is wrong. This does not capture the intuitions described above about the best interpretation of *Fish Tank*, i.e., that it is true in *Fish Tank* that Sara did *not* do a good thing. As we will see below, in order to arrive at such an interpretation, we we need to establish a communion behind a narrator's back.

Alternatively, the anaphoric subject of the hedge $\text{ACCEPT}_{s?}$ may bind to one of the fictional characters. However, in this particular text there are no textual clues that either of the salient available fictional characters (Sara or her father) is to be understood as uttering these evaluative words (out loud or silently in thought). We see none of the (sometimes subtle and ambiguous) textual and contextual clues that would license a free indirect discourse or protagonist projection interpretation here (see e.g., Eckardt (2014); Hinterwimmer (2017); Altshuler and Maier (2020b); Stokke (forthcoming-b); Abrusán (forthcoming)).

What we're left with is the option of accommodating a new fictional character, s', who is presumed to be offering this evaluation in a speech act: a fictional speaker/narrator responsible for telling the story, or at least this final part of it:[18]

(190)

x', y', u', v' s'	W
1:author(x') 2:fish.tank(y') 3: wrote(x',y') 4:killing.animals.is.wrong ... 5:Sara(u') 6:father(v',u') ...	
10:$\text{ACCEPT}_{s'}$ did.good(u')	
11:narrator(s')	
$\{5\dots11\} > 4 > \{1,2,3\}$	

After fictive closure on this workspace the reader considers it common ground between her and the author that there's a story called *Fish Tank* in which there is a girl named Sara who kills her father's fish. Moreover, in this story killing animals is morally wrong, as in the real world, and finally, this story is (partly) told from the perspective of a fictional narrator who claims that Sara did a good thing killing the fish. In other words, we started out interpreting the text on a par with the

---

[18]Altshuler and Maier (2020b) coin the term 'narrator accommodation' and argue that this is what causes the disruptive experience that is inherent in the phenomenon of imaginative resistance.

Heinlein story, i.e., as a third-person omniscient or rather impersonal narration, every statement to be taken at face value without the mediation of a personal narrator, and then switched to an interpretation along the lines of our Huckleberry Finn interpretation, i.e., as a first-person narration, all statements weighed against the available contextual and textual evidence and potentially treated as representing merely the point of view of the fictional character narrating the story.

Note that similar speaker accommodation (and hence a similar disruptive experience as in standard cases of imaginative resistance) may occur in the case of non-fictional discourse. Consider a case where I listen in on someone telling a story and don't realize that they are in fact reproducing someone else's speech. At some point they reach a part of the narrative that includes statements that are outrageously inconsistent with what I know of the speaker's own beliefs. I may then choose to accommodate an extra 'asserter' (separate from the actual speaker/author) who is committed to the truth of the relevant statements.

8.5 Conclusion

We can't always take a text at face value – so-called unreliable narrators may present a confused or misleading picture of the fictional world, and in cases of imaginative resistance readers refuse to accept parts of a text as true in the corresponding fictional world. We have proposed a way to model the interpretive processes that allow readers to extract fictional truths from such fictional narratives by combining a psychologistic version of the workspace account (where DRS's represent the hearer's conception of the common ground) with insights from belief revision theory. Our starting point has been that revision processes should apply uniformly across fiction and non-fiction when dealing with unreliable information sources.

Combining these two theoretical additions to the established DRT framework allowed us to describe precisely the various interpretation strategies readers can choose when interpreting different types of narratives. On our analysis, the 'epistemic entrenchment' of certain background assumptions relative to incoming information from the discourse or text, as well as the presence or absence of a personified speaker/narrator are the key factors in determining the kinds of readings available.

9 Conclusions

The central aim of this dissertation has been to model fictional discourse and related forms of discourse in terms of common ground updates. To this end I have attempted to combine insights from Matravers' theory of fiction interpretation and Stalnaker's common ground framework. Put very briefly, speaker and hearer first update a temporary workspace that is uniform for fiction and non-fiction and that only exists for the purpose and duration of a particular conversation. Secondly, the stable common ground is updated through either fictive or assertive closure.

When applying this basic framework to other types of discourse, I sometimes found that the initial account needed to be extended. For instance, I argued that in order to fit lies into the framework, the stable common ground could no longer be construed as belief-based. This requires either a moderate psychologistic turn (a departure from Stalnaker) or an acceptance-based stable common ground (a departure from Matravers since we no longer model discourse updates as leading to belief updates). Another example is metafictional discourse which, although it seems strongly related and in a sense even dependent on fictional discourse, is hard to model in the basic framework. It requires incorporating abstract objects or 'dot-objects' into our theory which complicates the analysis of parafictional discourse and fictive closure. Lastly, in order to model unreliable narration and inconsistent workspace updates we need to incorporate some kind of belief revision tools into the theory. Here we saw that holding on to both complete Matraversian uniformity of our primary engagement with fiction and non-fiction *and* Stalnakerian uniformity of how speaker and hearer engage with a discourse seems untenable because fiction can involve common ground revision whereas non-fiction cannot.

On the other hand, sometimes I found that the basic account provided neat analyses or explanations of certain phenomena that it was not designed to account for. In particular, the parafictional update at fictive closure leads us to an analysis of bald-faced lies as fictional statements. Moreover, it allows for a novel explanation of export of fictional truth (as based on analogical reasoning) that avoids the difficulties with existing accounts that maintain that fictions contain (indirect) assertions.

Looking ahead, it is important to keep in mind that Stalnaker's theoretical focus lies on common ground updates in *conversations*. Likewise, the workspace account focuses on explicit story telling (e.g., campfire stories or novels) as its prime example of fiction practices; its main aim is to explain linguistic behaviour involved

in engaging with fiction. However, to the extent that one can also analyse nonverbal communication (e.g., music, dance, paintings, pictures, movies) as proposals to update a common ground between 'speaker' and 'hearer', the workspace account should be extendable to nonverbal media.

10 Publieke samenvatting

In het algemeen vertrouw ik erop dat anderen mij de waarheid vertellen. Wanneer ik een vreemdeling bij de bushalte vraag of lijn 4 hier stopt, dan verwacht ik dat de vreemdeling (onbewust) het Maxime van Kwaliteit (Grice (1989)) volgt. Dat wil zeggen, ik verwacht dat mijn conversationele partner zal proberen iets waars te zeggen. Dit vertrouwen is essentieel; als er geen vertrouwen is binnen een samenleving dat (de meeste) mensen deelnemen aan coöperatieve informatie uitwisseling, dan lopen we het risico dat onze gehele taalpraktijk zijn nut verliest. Waarom zou ik überhaupt proberen de vreemdeling over de bus te vragen als ik er niet op kan vertrouwen dat hij iets waars zal proberen te zeggen?

Tegelijkertijd vinden mensen het simpelweg geweldig om anderen het Maxime van Kwaliteit te zien breken of zelfs om het maxime zelf te breken. We vinden het heerlijk om overduidelijk onware verslagen te lezen of te bekijken over onwaarschijnlijke romances in het Engeland van de negentiende eeuw, vuurspuwende draken, pratende dieren, heroïsche aliens en futuristische samenlevingen. We vullen niet alleen onze boekenkasten, slaapkamermuren en hoofden met verhalen en beelden waarvan we weten dat ze onware 'feiten' rapporteren, we bediscussiëren deze onwaarheden ook graag uitgebreid. Middelbare scholen verplichten hun leerlingen om de beroemdste onware verhalen van hun samenleving te bestuderen (bijvoorbeeld *Hamlet* of *Van den Vos Reynaerde*) en mensen creëren samen uitgebreide online encyclopedieën die rapporteren over hun favoriete verzameling onwaarheden (bijvoorbeeld *One Wiki to Rule Them All* of *Wookieepedia*). Mensen kunnen juist genieten van dit overduidelijk breken van het Maxime van Kwaliteit omdat het voor iedereen duidelijk is dat de relevante verhalen onwaar zijn. Dit is een van de belangrijkste kenmerken van fictie; niemand wordt misleid om iets onwaars te geloven en hierdoor kunnen we ons algehele vertrouwen in conversationele partners behouden.

De focus van deze dissertatie ligt op het modelleren van verschillende manieren waarop we taal gebruiken wanneer we ons bezighouden met fictionele verhalen. Dit algemene doel splitst zich op in een aantal puzzels omtrent verschillende soorten uitspraken (bijvoorbeeld 'gewone' non-fictionele asserties, fictionele uitspraken, leugens, asserties over fictie, etc.). In deze dissertatie ontwikkel ik een coherente semantische analyse van deze verschillende puzzels: de 'workspace account'. Deze theorie is een extensie van Stalnakers (1970, 1978, 1984, 2002) beroemde pragmatische 'common ground' framework. In dit framework worden asserties

gemodelleerd als voorstellen om de 'common ground' (de verzameling gedeelde aannames) tussen conversationele partners uit te breiden.

Hoofdstuk 2 introduceert het concept 'common ground' en beargumenteert dat standaard formalisaties van deze notie hem ontoepasbaar maken op bijvoorbeeld de conversatie tussen een schrijver en zijn lezers wanneer hij deze niet persoonlijk kent. Ik ontwikkel en vergelijk twee alternatieve formalisaties die wel toepasbaar zijn.

Hoofdstuk 3 introduceert de centrale puzzel in de studie naar de semantiek van fictie: hoe onderscheiden we 'normale' non-fictionele uitspraken van fictionele uitspraken? Wat gebeurde er toen Tolkien (191) opschreef in *The Hobbit* en hoe verschilt dat van zijn non-fictionele uitspraken?

(191)   In een hol onder de grond woonde een hobbit.

Ik formaliseer en bekritiseer bestaande voorstellen om fictie te modelleren in het common ground framework: fictionele uitspraken zijn voorstellen om aparte 'unofficial common grounds' uit te breiden (Eckardt (2014); Stokke (2013, 2018)).

In hoofdstuk 4 ontwikkel ik een alternatieve theorie (de workspace account) die is geïnspireerd door de theorie van Matravers (2014) over fictie interpretatie en Lewis' (1978) analyse van de 'In verhaal s, ϕ'-operator.

In hoofdstuk 5 evalueer ik verschillende manieren om leugens te incorporeren in een common ground framework. Ik beargumenteer dat zogenaamde 'bald-faced lies' (leugens waarbij er geen intentie is om de hoorder te misleiden) het beste gemodelleerd kunnen worden als fictionele uitspraken.

In hoofdstuk 6 introduceer ik zogenaamde 'metafictionele' uitspraken. Dit zijn uitspraken waarmee iemand praat over een fictionele entiteit *als fictionele entiteit* zoals in (192):

(192)   Frodo is verzonnen door Tolkien.

Ik vergelijk vier verschillende strategieën om deze uitspraken te incorporeren in de workspace account.

In hoofdstuk 7 beargumenteer ik dat de operatoren 'In story s, ϕ' ('In verhaal s, ϕ') en 'According to story s, ϕ' ('Volgens verhaal s, ϕ') – die vaak een uniforme analyse krijgen – eigenlijk op verschillende manieren gebruikt worden. Ik stel twee aparte semantische analyses voor voor deze twee operatoren die dit kunnen verklaren.

Hoofdstuk 8 is grotendeels samen geschreven met Dr. Emar Maier en ontwikkelt een extensie van de workspace account waarin uitspraken van onbetrouwbare vertellers worden gemodelleerd. Dit zijn uitspraken in een fictionele tekst waarbij we niet zeker kunnen zijn dat wat er staat ook daadwerkelijk waar is in de fictionele wereld.

Bibliography

Abreu Zavaleta, M. (2019). Weak speech reports. *Philosophical Studies*, 176(8):2139–2166.

Abrusán, M. (forthcoming). Computing perspective shift in narratives. In Maier, E. and Stokke, A., editors, *The Language of Fiction*. Oxford University Press, Oxford.

Abusch, D. (1997). Sequence of tense and temporal de re. *Linguistics and Philosophy*, 20(1):1–50.

Akkoyunlu, E. A., Ekanadham, K., and Huber, R. V. (1975). Some constraints and tradeoffs in the design of network communications. *SIGOPS Operating Systems Review*, 9(5):67–74.

Alchourron, C. E., Gardenfors, P., and Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50(2):510–530.

Altshuler, D. and Maier, E. (2020a). Coping with imaginative resistance. manuscript. https://philpapers.org/rec/ALTCWI.

Altshuler, D. and Maier, E. (2020b). Death on the freeway: Imaginative resistance as narrator accommodation. In Bhatt, R., Frana, I., and Menéndez-Benito, P., editors, *Making Worlds Accessible: Festschrift for Angelika Kratzer*. Amherst.

Anand, P. and Hacquard, V. (2014). Factivity, belief and discourse. In Crnič, L. and Sauerland, U., editors, *The Art and Craft of Semantics: A Festschrift for Irene Heim*, volume 1, pages 69–90. MIT Working Papers in Linguistics.

Anand, P. and Korotkova, N. (2019). Speech reports. lecture notes. http://esslli2019.folli.info/courses/speech-reports/.

Andow, J. (2019). Why don't we trust moral testimony? *Mind & Language*, 35(4):456–474.

Asher, N. (2011). *Lexical Meaning in Context: A Web of Words*. Cambridge University Press, Cambridge.

Bibliography

Augustine, S. (1952 [395]). Lying. In Deferrari, R. J., editor, *Treatises on Various Subjects*, volume 16, pages 45–110. Catholic University of America Press, Washington, D.C.

Badura, C. and Berto, F. (2019). Truth in fiction, impossible worlds, and belief revision. *Australasian Journal of Philosophy*, 97(1):178–193.

Bartha, P. (2010). *By Parallel Reasoning: The Construction and Evaluation of Analogical Arguments*. Oxford University Press, Oxford.

Barwise, J. (1988). Three views of common knowledge. In *Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning about Knowledge*, TARK '88, pages 365–379, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Bary, C. and Maier, E. (2020). The landscape of speech reporting. manuscript. https://ling.auf.net/lingbuzz/003526.

Bermúdez, J. L. (2003). *Thinking Without Words*. Oxford University Press, Oxford.

Bjurman Pautz, A. (2008). Fictional coreference as a problem for the pretense theory. *Philosophical Studies*, 141(2):147–156.

Booth, W. (1961). *The Rhetoric of Fiction*. UCP, Chicago.

Braithwaite, R. B. (1932). The nature of believing. *Proceedings of the Aristotelian society*, 33(1):129–146.

Brandenburger, A. and Dekel, E. (1987). Common knowledge with probability 1. *Journal of Mathematical Economics*, 16(3):237–245.

Brasoveanu, A. and Farkas, D. (2007). Say reports, assertion events and meaning dimensions. In Alboiu, G., Avram, A., Avram, L., and Isac, D., editors, *Pitar Mos: A Building with a View. Papers in Honour of Alexandra Cornilescu*, pages 175–196. Editura Universităţii din Bucureşt, Bucharest.

Cargile, J. (1970). A note on "iterated knowings". *Analysis*, 30(5):151–155.

Carlson, G. N. (1982). Generic terms and generic sentences. *Journal of Philosophical Logic*, 11(2):145–181.

Carson, T. L. (2006). The definition of lying. *Noûs*, 40(2):284–306.

Clark, H. H. and Marshall, C. R. (1981). Definite reference and mutual knowledge. In Joshi, A. K., Webber, B. L., and Sag, I. A., editors, *Elements of Discourse Understanding*, pages 10–63. Cambridge University Press, Cambridge.

Cohen, D. and Yemini, Y. (1979). Protocols for dating coordination. In *Proceedings of the fourth Berkeley conference on distributed data management and computer networks*, page 179–188. Lawrence Berkeley Laboratory.

Collins, J. (2019). The diversity of fiction and copredication: An accommodation problem. *Erkenntnis*.

Condoravdi, C. (2002). Temporal interpretation of modals. modals for the present and for the past. In Beaver, D., Kaufmann, S., Clark, B., and Casillas, L., editors, *The Construction of Meaning*, pages 59–88. CSLI, Stanford.

Cooper, R. (2007). Copredication, dynamic generalized quantification and lexical innovation by coercion. In Bouillon, P., Danlos, L., and Kanzaki, K., editors, *Fourth International Workshop on Generative Approaches to the Lexicon*, pages 143–184.

Cresswell, M. J. (1985). *Structured meanings: The Semantics of Propositional Attitudes*. MIT Press, Cambridge, MA, USA.

Cresswell, M. J. and Von Stechow, A. (1982). "De re" belief generalized. *Linguistics and Philosophy*, 5(4):503–535.

Currie, G. (1990). *The Nature of Fiction*. Cambridge University Press, Cambridge.

David, M. (2015). The fragmented mind: Agency, belief, rationality. https://philosophie-gewi.uni-graz.at/en/theoretical-philosophy/projects/the-fragmented-mind/.

Dynel, M. (2011). A web of deceit: A neo-gricean view on types of verbal deception. *International Review of Pragmatics*, 3(2):139–167.

Eckardt, R. (2014). *The Semantics of Free Indirect Discourse*. Brill Publishers, Leiden.

Elbourne, P. (2005). *Situations and Individuals*. MIT Press, Cambridge, MA, USA.

Enç, M. (1996). Tense and modality. In Lappin, S., editor, *The Handbook of Contemporary Semantic Theory*, pages 345–358. Blackwell Publishers Ltd., Oxford.

Evans, G. (1977). Pronouns, quantifiers and relative clauses. *Canadian Journal of Philosophy*, 8(3):467–536.

Evans, G. (1982). *The Varieties of Reference*, volume 10. Oxford University Press, Oxford.

Everett, A. (2013). *The Nonexistent*. Oxford University Press, Oxford.

Bibliography

Fagin, R., Halpern, J., Moses, Y., and Vardi, M. (1995). *Reasoning About Knowledge*. MIT Press, Cambridge, MA, USA.

Fagin, R., Halpern, J. Y., Moses, Y., and Vardi, M. Y. (1999). Common knowledge revisited. *Annals of Pure and Applied Logic*, 96(1):89–105.

Fallis, D. (2009). What is lying? *Journal of Philosophy*, 106(1):29–56.

Franzén, N. (forthcoming). Truth in fiction: In defense of the reality principle. In Maier, E. and Stokke, A., editors, *The Language of Fiction*. Oxford University Press, Oxford.

Friedell, M. F. (1969). On the structure of shared awareness. *Behavioral Science*, 14(1):28–39.

Friend, S. (2008). Imagining fact and fiction. In Stock, K. and Thomson-Jones, K., editors, *New Waves in Aesthetics*, pages 150–169. Palgrave Macmillan UK, London.

Friend, S. (2011a). Fictive utterance and imagining ii. *Supplementary Proceedings of the Aristotelian Society*, 85:163–180.

Friend, S. (2011b). The great beetle debate: a study in imagining with names. *Philosophical Studies*, 153(2):183–211.

Friend, S. (2012). Fiction as a genre. *Proceedings of the Aristotelian Society*, 112(2):179–209.

Friend, S. (2017). The real foundation of fictional worlds. *Australasian Journal of Philosophy*, 95(1):29–42.

Gans, C. (1970). How snakes move. *Scientific American*, 222(6):82–99.

García-Carpintero, M. (2013). Norms of fiction-making. *British Journal of Aesthetics*, 53(3):339–357.

Geach, P. T. (1967). Intentional identity. *The Journal of Philosophy*, 64(20):627–632.

Gendler, T. S. (2000). The puzzle of imaginative resistance. *Journal of Philosophy*, 97(2):55–81.

Geurts, B. (1997). Communication as commitment sharing: speech acts, implicatures, common ground. *Theoretical linguistics*, 45(1-2):1–30.

Geurts, B. (1999). *Presuppositions and Pronouns*. Elsevier, Amsterdam.

Geurts, B. (2020). Common ground as a normative condition. manuscript. https://www.academia.edu/43209457/Common_ground_as_a_normative_ condition.

Geurts, B., Beaver, D. I., and Maier, E. (2016). Discourse Representation Theory. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2016 edition.

Gotham, M. (2017). Composing criteria of individuation in copredication. *Journal of Semantics*, 34:333–371.

Gray, J. N. (1978). Notes on data base operating systems. In Bayer, R., Graham, R. M., and Seegmüller, G., editors, *Operating Systems: An Advanced Course*, pages 393–481. Springer, Berlin.

Green, M. (2017). Conversation and common ground. *Philosophical Studies*, 174:1587–1604.

Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press, Cambridge, MA, USA.

Groenendijk, J. and Stokhof, M. (1991). Dynamic predicate logic. *Linguistics and Philosophy*, 14(1):39–100.

Grove, A. (1988). Two modellings for theory change. *Journal of Philosophical Logic*, 17(2):157–170.

Gärdenfors, P. (1988). *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press, Cambridge, MA, USA.

Gärdenfors, P. and Makinson, D. (1988). Revisions of knowledge systems using epistemic entrenchment. In *Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning About Knowledge*, TARK '88, pages 83–95, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Halpern, J. Y. and Moses, Y. (1990). Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37(3):549–587.

Hamm, F., Kamp, H., and Van Lambalgen, M. (2006). There is no opposition between formal and cognitive semantics. *Theoretical Linguistics*, 32(1):1–40.

Hansson, S. O. (1994). Taking belief bases seriously. In Prawitz, D. and Westerståhl, D., editors, *Logic and Philosophy of Science in Uppsala: Papers from the 9th International Congress of Logic, Methodology and Philosophy of Science*, pages 13–28, Dordrecht. Springer.

Bibliography

Hansson, S. O. (1998). Revision of belief sets and belief bases. In Dubois, D. and Prade, H., editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems Volume 3: Belief Change*, pages 17–75. Springer, Dordrecht.

Harris, D. W. (forthcoming). Intentionalism and bald-faced lies. *Inquiry*.

Heal, J. (1978). Common knowledge. *The Philosophical Quarterly*, 28(111):116–131.

Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, UMass Amherst.

Heim, I. (1990). E-type pronouns and donkey anaphora. *Linguistics and Philosophy*, 13:137–178.

Hesse, M. B. (1966). *Models and Analogies in Science*. University of Notre Dame Press, Notre Dame.

Hinterwimmer, S. (2017). Prominent protagonists. *Journal of Pragmatics*, 154:79–91.

Hintikka, J. (1962). *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press, Ithaca.

Hunter, D. (1981). Reference and meinongian objects. *Grazer Philosophische Studien*, 14:23–36.

Hunter, J. (2013). Structured contexts and anaphoric dependencies. *Philosophical Studies*, 168:35–58.

Ichino, A. and Currie, G. (2017). Truth and trust in fiction. In Sullivan-Bissett, E., Bradley, H., and Noordhof, P., editors, *Art and Belief*, pages 63–82. Oxford University Press, Oxford.

Jeshion, R. (2010). Singular thought: Acquaintance, semantic instrumentalism, and cognitivism. In Jeshion, R., editor, *New Essays on Singular Thought*, pages 105–141. Oxford University Press, Oxford.

Kamp, H. (1981). A theory of truth and semantic representation. In Groenendijk, J. A. G., Janssen, T. M. V., and Stokhof, M. B. J., editors, *Formal Methods in the Study of Language, Part 1*, pages 277–322. Blackwell Publishers Ltd, Oxford.

Kamp, H. (1990). Prolegomena to a structural account of belief and other attitudes. In Anderson, A. and Owens, J., editors, *Propositional Attitudes: The Role of Content in Logic, Language, and Mind*, pages 27–90. CSLI, Stanford.

Kamp, H. (2015). Using proper names as intermediaries between labelled entity representations. *Erkenntnis*, 80(2):263–312.

Kamp, H. (2018). Entity representations and articulated contexts: An exploration of the semantics and pragmatics of definite noun phrases. manuscript. https://www.ims.uni-stuttgart.de/archiv/kamp/files/2018.kamp.articulated.contexts.pdf.

Kamp, H. (forthcoming). Sharing real and fictional reference. In Maier, E. and Stokke, A., editors, *The Language of Fiction*. Oxford University Press, Oxford.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic: an Introduction to Modeltheoretic Semantics in Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer, Dordrecht.

Kaplan, D. (1968). Quantifying in. *Synthese*, 19(1-2):178–214.

Kaufmann, M. and Kaufmann, S. (2020). Talking about sources. manuscript. http://magdalena-kaufmann.uconn.edu/papers/givenPREFINAL.pdf.

Keiser, J. (2016). Bald-faced lies: how to make a move in a language game without making a move in a conversation. *Philosophical Studies*, 173(2):461–477.

Keiser, J. (forthcoming). The limits of acceptance. *Inquiry*.

Kim, H., Kneer, M., and Stuart, M. T. (2018). The context-dependence of imaginative resistance. In Cova, F. and Réhault, S., editors, *Advances in Experimental Philosophy of Aesthetics*, pages 194–224. Bloomsbury, London.

Kissine, M. (2008). Why will is not a modal. *Natural Language Semantics*, 16:129–155.

Klauk, T. (2014). Zalta on encoding fictional properties. *Journal of Literary Theory*, 8(2):234–256.

Krawczyk, E. (2012). *Inferred propositions and the expression of the evidence relation in natural language: Evidentiality in Central Alaskan Yup'ik Eskimo and English*. PhD thesis, Georgetown University.

Kripke, S. (1980). *Naming and Necessity*. Harvard University Press, Cambridge, MA, USA.

Kripke, S. (2011). Vacuous names and fictional entities. In *Philosophical Troubles*, pages 52–74. Oxford University Press, Oxford.

Kripke, S. (2013 [1973]). *Reference and Existence. The John Locke Lectures for 1973*. Oxford University Press, Oxford.

Lackey, J. (2013). Lies and deception: an unhappy divorce. *Analysis*, 73(2):236–248.

Bibliography

Lamarque, P. (1990). Reasoning to what is true in fiction. *Argumentation*, 4(3):333–346.

Leland, P. R. (2015). Rational responsibility and the assertoric character of bald-faced lies. *Analysis*, 75(4):550–554.

Lewis, D. (1969). *Convention: A Philosophical Study*. Harvard University Press, Cambridge, MA, USA.

Lewis, D. (1973). *Counterfactuals*. Blackwell Publishers Ltd., Oxford.

Lewis, D. (1978). Truth in fiction. *American Philosophical Quarterly*, 15(1):37–46.

Lewis, D. (1979). Attitudes de dicto and de se. *The Philosophical Review*, 88(4):513–543.

Liao, S. and Gendler, T. S. (2016). The problem of imaginative resistance. In Carroll, N. and Gibson, J., editors, *The Routledge Companion to Philosophy of Literature*, pages 405–418. Routledge, London.

Luo, Z. (2012). Formal semantics in modern type theories with coercive subtyping. *Linguistics and Philosophy*, 35(6):491–513.

Maier, E. (2009a). Iterated de re: A new puzzle for the relational report semantics. In Riester, A. and Solstad, T., editors, *Proceedings of Sinn und Bedeutung 13*, pages 347–355.

Maier, E. (2009b). Proper names and indexicals trigger rigid presuppositions. *Journal of Semantics*, 26(3):253–315.

Maier, E. (2017). Fictional names in psychologistic semantics. *Theoretical Linguistics*, 43(1-2):1–42.

Maier, E. (2019). On the exceptionality of reported speech. *Linguistic Typology*, 23(1):197–205.

Maier, E. and Semeijn, M. (forthcoming). Extracting fictional truth from unreliable sources. In Maier, E. and Stokke, A., editors, *The Language of Fiction*. Oxford University Press, Oxford.

Maitra, I. (2018). Lying, acting, and asserting. In Michaelson, E. and Stokke, A., editors, *Lying: Language, Knowledge, Ethics, and Politics*, pages 73–94. Oxford University Press, Oxford.

Mally, E. (1912). *Gegenstandstheoretische Grundlagen der Logik und Logistik.* Ergänzungsheft zu Bd. 148 der Zeitschrift für Philosophie und Philosophische Kritik. Barth, Leipzig.

Marques, T. (2020). Disagreement with a bald-faced liar. *Ratio*, 33(4):255–268.

Matravers, D. (2014). *Fiction and Narrative.* Oxford University Press, Oxford.

Meggle, G. (2003). Common belief and common knowledge. In Sintonen, M., Ylikoski, P., and Miller, K., editors, *Realism in Action: Essays in the Philosophy of the Social Sciences*, pages 251–258. Springer, Dordrecht.

Meibauer, J. (2014). Bald-faced lies as acts of verbal aggression. *Journal of Language Aggression and Conflict*, 2:127–150.

Meskin, A. and Weinberg, J. M. (2006). Imagine that! In Kieran, M., editor, *Contemporary Discussions in Aesthetics and the Philosophy of Art*, pages 222–235. Blackwell Publishers Ltd., Oxford.

Meyer, J. J. C. and van der Hoek, W. (1995). *Epistemic Logic for AI and Computer Science.* Cambridge Tracts in Theoretical Computer Science. Cambridge University Press.

Monderer, D. and Samet, D. (1989). Approximating common knowledge with common beliefs. *Games and Economic Behavior*, 1(2):170–190.

Murday, B. (2010). Two-dimensionalism and fictional names. In Lihoreau, F., editor, *Truth in Fiction*, pages 43–76. De Gruyter, Berlin, Boston.

Myhill, J. (1963). An alternative to the method of extension and intension. In Schilpp, P. A., editor, *The philosophy of Rudolf Carnap*, pages 299–310. Open Court Publishing Company, Chicago, IL, USA.

Nebel, B. (1998). How hard is it to revise a belief base? In Dubois, D. and Prade, H., editors, *Belief Change*, Handbook of Defeasible Reasoning and Uncertainty Management Systems, pages 77–145. Springer, Dordrecht.

Nichols, S. and Stich, S. (2000). A cognitive theory of pretense. *Cognition*, 74:115–47.

Ninan, D. (2017). Names in fiction. *Theoretical Linguistics*, 43(1-2).

Norton, J. D. (2021). The material theory of induction. manuscript. https://www.pitt.edu/j̃dnorton/papers/material_theory/Material_Induction_ March_14_2021.pdf.

Bibliography

Panangaden, P. and Taylor, K. (1998). Concurrent common knowledge: Defining agreement for asynchronous systems. *Distributed Computing*, 6:73–93.

Prior, A. (1967). *Past, Present and Future*. Oxford University Press, Oxford.

Pustejovsky, J. (1995). *The Generative Lexicon*. MIT Press, Cambridge, MA, USA.

Quine, W. V. O. (1956). Quantifiers and propositional attitudes. *Journal of Philosophy*, 53(5):177–187.

Recanati, F. (2009). Singular thought: In defense of acquaintance. In Jeshion, R., editor, *New essays on singular thought*, pages 141–188. Oxford University Press, Oxford.

Recanati, F. (2012). *Mental Files*. Oxford University Press, Oxford.

Recanati, F. (2018). Fictional, metafictional, parafictional. *Proceedings of the Aristotelian Society*, 118(1):25–54.

Recanati, F. (forthcoming-a). Entertaining as simulation. In Mras, G. and Schmitz, M., editors, *Force, Content and the Unity of the Proposition*. Routledge, London.

Recanati, F. (forthcoming-b). Fictional reference as simulation. In Maier, E. and Stokke, A., editors, *The Language of Fiction*. Oxford University Press, Oxford.

Reid, T. (2011 [1785]). *Essays on the Intellectual Powers of Man*. Cambridge Library Collection. Cambridge University Press, Cambridge.

Rendsvig, R. and Symons, J. (2019). Epistemic logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2019 edition.

Ross, J. (2012). *The Semantics of Media*. Studies in Linguistics and Philosophy. Springer, Dordrecht.

Russell, B. (1905). On denoting. *Mind*, 14(56):479–493.

Ryan, M. (1980). Fiction, non-factuals, and the principle of minimal departure. *Topics in Catalysis*, 9(4):403–422.

Ryan, M. (1991). *Possible Worlds, Artificial Intelligence, and Narrative Theory*. Indiana University Press, Bloomington, IN, USA.

Sainsbury, R. M. (2014). Fictional worlds and fiction operators. In García-Carpintero, M. and Martí, G., editors, *Empty Representations*, pages 277–289. Oxford University Press, Oxford.

Salkie, R. (2010). Will: tense or modal or both? *English Language & Linguistics*, 14(2):187–215.

Saul, J. M. (2012). Lying, misleading, and what is said: An exploration in philosophy of language and ethics. *Analysis*, 74(1):179–180.

Schiffer, S. (1972). *Meaning*. Clarendon Press, Oxford.

Schiffer, S. (2003). *The Things We Mean*. Clarendon Press, Oxford.

Searle, J. R. (1975). The logical states of fictional discourse. *New Literary History*, 6(2):319–332.

Semeijn, M. (2017). A Stalnakerian analysis of metafictive statements. In *Proceedings of the 21st Amsterdam Colloquium*. ILLC.

Semeijn, M. (2019). Interacting with fictions: The role of pretend play in Theory of Mind acquisition. *Review of Philosophy and Psychology*, 10(1):113–132.

Semeijn, M. and Bimpikou, S. (2019). Parafictional tense. poster for SPE11. https://merelsemeijn.files.wordpress.com/2019/10/poster-final.pdf.

Semeijn, M. and Zalta, E. N. (2021). Revisiting the 'wrong kind of object' problem. *Organon F*, 28(1):168–197.

Soames, S. (1985). Lost innocence. *Linguistics and Philosophy*, 8(1):59–71.

Sorensen, R. (2007). Bald-faced lies! Lying without the intent to deceive. *Pacific Philosophical Quarterly*, 88(2):251–264.

Stalnaker, R. C. (1968). A theory of conditionals. In Rescher, N., editor, *Studies in Logical Theory: American Philosophical Quarterly Monograph Series 2*, volume 6, page 98–112. Blackwell Publishers Ltd., Oxford.

Stalnaker, R. C. (1970). Pragmatics. *Synthese*, 22(1-2):272–289.

Stalnaker, R. C. (1978). Assertion. In Cole, P., editor, *Pragmatics*, volume 9, pages 315–32. New York Academic Press, New York.

Stalnaker, R. C. (1984). *Inquiry*. MIT Press, Cambridge, MA, USA.

Stalnaker, R. C. (2002). Common ground. *Linguistics and Philosophy*, 25(5-6):701–721.

Stinchcombe, M. (1988). Approximate common knowledge. mimeograph. San Diego: University of California.

Bibliography

Stock, K. (2011). Fictive utterance and imagining. *Supplementary Proceedings of the Aristotelian Society*, 85:145–161.

Stokke, A. (2013). Lying and asserting. *Journal of Philosophy*, 110(1):33–60.

Stokke, A. (2018). *Lying and Insincerity*. Oxford University Press, Oxford.

Stokke, A. (forthcoming-a). Fictional names and individual concepts. *Synthese*.

Stokke, A. (forthcoming-b). Protagonist projection, character focus, and mixed quotation. In Maier, E. and Stokke, A., editors, *The Language of Fiction*. Oxford University Press, Oxford.

Sæbø, K. J. (2013). Reports of specific indefinites. *Journal of Semantics*, 30(3):267–314.

Toshyuki, O. and Sharvit, Y. (2012). Embedded tenses. In Binnick, R. I., editor, *The Oxford Handbook of Tense and Aspect*. Oxford University Press, Oxford.

van der Sandt, R. (1992). Presupposition projection as anaphora resolution. *Journal of Semantics*, 9(4):333–377.

van Ditmarsch, H., Hendriks, P., and Verbrugge, R. (2020). Editors' review and introduction: Lying in logic, language, and cognition. *Topics in Cognitive Science*, 12(2):466–484.

van Ditmarsch, H., van Eijck, J., and Verbrugge, R. (2009). Common knowledge and common belief. In van Eijck, J. and Verbrugge, R., editors, *Discourses on Social Software, Texts in Logic and Games*, pages 99–122. Amsterdam University Press, Amsterdam.

van Inwagen, P. (1977). Creatures of fiction. *American Philosophical Quarterly*, 14(4):299–308.

van Inwagen, P. (2000). Quantification and fictional discourse. In Everett, A. and Hofweber, T., editors, *Empty Names, Fiction and the Puzzles of Non-Existence*, pages 235–47. CSLI, Stanford.

Vanderschraaf, P. and Sillari, G. (2014). Common knowledge. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2014 edition.

Voltolini, A. (2019). Varieties of fiction operators: Methods and applications. In Capone, A., Carapezza, M., and Lo Piparo, F., editors, *Further Advances in Pragmatics and Philosophy: Part 2 Theories and Applications*, pages 199–210. Springer.

von Fintel, K. and Heim, I. (2011). *Intensional Semantics*. MIT. unpublished textbook.

von Stechow, A. and Zimmermann, E. (2005). A problem for a compositional treatment of de re attitudes. In Carlson, G. and Pelletier, F., editors, *Reference and Quantification: The Partee Effect*, pages 207–228. CSLI, Stanford.

Walton, K. L. (1990). *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Harvard University Press, Cambridge, MA, USA.

Weatherson, B. (2004). Morality, fiction, and possibility. *Philosopher's Imprint*, 4(3).

Wieland, N. (forthcoming). Metalinguistic acts in fiction. In Maier, E. and Stokke, A., editors, *The Language of Fiction*. Oxford University Press, Oxford.

Williams, B. (2002). *Truth and Truthfulness: An Essay in Genealogy*. Princeton University Press, Princeton.

Williams, M. (1994). On the logic of theory base change. In MacNish, C., Pearce, D., and Pereira, L. M., editors, *Logics in Artificial Intelligence*, Lecture Notes in Computer Science, pages 86–105, Berlin. Springer.

Yablo, S. (2002). Coulda, woulda, shoulda. In Gendler, T. S. and Hawthorne, J., editors, *Conceivability and Possibility*, pages 441–492. Oxford University Press, Oxford.

Zalta, E. N. (1983). *Abstract Objects: An Introduction to Axiomatic Metaphysics*. Springer, Berlin.

Zalta, E. N. (1988). *Intensional Logic and the Metaphysics of Intentionality*. MIT Press, Cambridge, MA, USA.

Zalta, E. N. (2000). The road between pretense theory and abstract object theory. In Hofweber, T. and Everett, A., editors, *Empty Names, Fiction, and the Puzzles of Non-Existence*, pages 117–147. CSLI, Stanford.

Zalta, E. N. (2003 [1987]). Erzählung als Taufe des Helden: Wie man auf fiktionale Objekte Bezug nimmt. *Zeitschrift für Semiotik*, 9(1–2):85–95. Translated (Arnold Günther) as Referring to fictional characters in *Dialectica*, 2(2003):243–254.

Zucchi, A. (forthcoming). On the generation of content. In Maier, E. and Stokke, A., editors, *The Language of Fiction*. Oxford University Press, Oxford.

Zucchi, S. (2001). Tense in fiction. In Cecchetto, C., Chierchia, G., and Guasti, M. T., editors, *Semantic Interfaces: Reference, Anaphora and Aspect*. CSLI, Stanford.

Zucchi, S. (2017). Games of make-believe and factual information. *Theoretical Linguistics*, 43(1-2).

Acknowledgements

I feel extremely lucky to have had the opportunity to focus on philosophical research and work on this dissertation for the last four years. My time as a PhD student has played a very important role in my personal development. I would like to take this opportunity to thank everyone who has contributed to this.

First and foremost, I would like to thank my absolutely excellent supervisor Emar Maier. Emar is a role model. Not only is he an extremely talented and serious researcher, he is also deeply invested in the well-being of his PhD students and exhibits a seemingly never-ending readiness to have impromptu meetings to discuss new worries, issues, questions and plans.

I would like to thank my colleague Sofia Bimpikou for being an awesome friend, making conferences and summer schools extra fun, inspiring discussions and fruitful collaboration.

I would like to thank everyone at the faculty of philosophy for creating a friendly and supportive atmosphere where young philosophers can thrive. In particular, I would like to thank the members of the semantics of fiction reading group for high quality discussions and my friends Cesar Reigosa Soler, Crystel Hajjar, Grzegorz Gaszczyk, Kritika Maheshwari, Michael Gregory, Sanne Hupkes, Stefan Sleeuw, Stipe Pandžić and Yuri David Santos.

I would like to thank the Semantics and Cognition group at the University of Groningen for allowing me to join their meetings and get a taste of the world outside of philosophy. I would also like to thank the people at the Institut Jean Nicod in Paris for the warm welcome I received during my research visit there.

Finally, I would like to thank my family and friends for their tremendous support, particularly at the final stages of the project. Especially I would like to thank my partner Jochem Dijkstra; my parents Jaap Semeijn and Marjolein Semeijn Visser; my brothers Koen Semeijn and Rik Semeijn; and my close friends Alissa Mali, Claire Wagemans, Justin Warners, Leandra du Pau, Sophie Starrenburg and Vanja Sarić.