

University of Groningen

Fiction and common ground

Semeijn, Merel

DOI:
[10.33612/diss.177806543](https://doi.org/10.33612/diss.177806543)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Semeijn, M. (2021). *Fiction and common ground: a workspace account*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.177806543>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

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 forthcoming 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

We would like to thank Andreas Stokke, the audience at the *Language of Fiction* workshop at Uppsala University, the audience at the *Linguistic Perspectives on Perspective-Taking in Narrative Discourse* workshop at the University of Wuppertal and the audience at the *Multiple Perspectives* workshop at Radboud University for valuable comments, questions, and discussions.

¹This formulation has been adapted from [Badura and Berto \(2019\)](#).

²Though 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

8 Unreliable narration and imaginative resistance

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 H₂O, 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

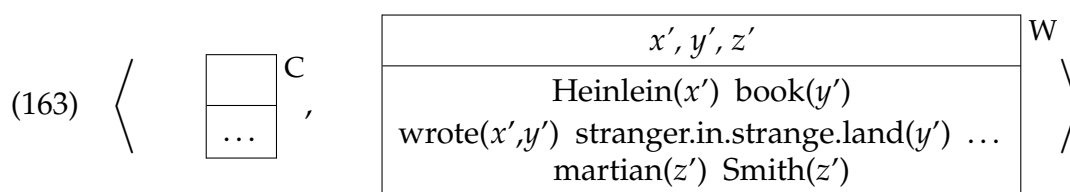
Separate from issues concerning unreliable narrators and imaginative resistance, there is in fact already a basic need for some kind of revision mechanism in the workspace account (see sections 3.3.1 and 4.4). This is because fictional stories often introduce content that is inconsistent with our shared 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*:

$$(161) \quad \left\langle \begin{array}{|c|} \hline x, y \\ \hline \text{Heinlein}(x) \text{ book}(y) \\ \text{wrote}(x,y) \text{ stranger.in.strange.land}(y) \dots \\ \hline \end{array} \right\rangle^{\mathcal{C}, \emptyset}$$

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.

8 Unreliable narration and imaginative resistance



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

³In a similar way fictive opening with genre conventions (e.g., that in fairytales, dragons breathe fire) can lead to inconsistencies in the workspace.

⁴Similarly, 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 $\neg p$ 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 $\neg p$ in the first place and hence that $\neg p$ 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 $\neg p$ 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 $\neg p$. 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 Matravvers' 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 Matravvers' theory per se because it entails a difference in the *production* of fictional versus non-fictional discourse, not in the *interpretation* of the discourse. Matravvers' theory really only predicts uniformity of the reader's interpretation of fictional and non-fictional discourse. In line with Matravvers, 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 psychologicistic 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

⁵By contrast, common ground revision that is prompted by for instance the shared perception of some event that contradicts previous shared background seems possible.

8 *Unreliable narration and imaginative resistance*

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 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

⁶With 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.

⁷In 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

8 Unreliable narration and imaginative resistance

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

⁸Consistency 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.

8.3 Fiction updates and belief revision

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:

$$(164) \quad K \div \phi = \{\psi \in K \mid \phi < (\phi \vee \psi), \text{ or } \phi \text{ is a tautology}\}$$

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

8 Unreliable narration and imaginative resistance

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:

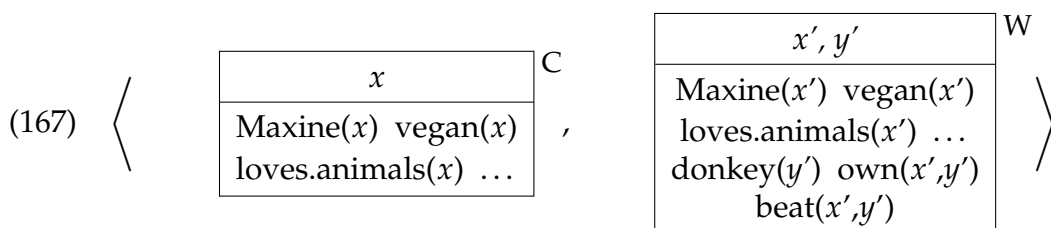
$$(165) \quad (K \cup \{p\}) \div \perp$$

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.



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-

8.3 Fiction updates and belief revision

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.)

⁹We 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.

¹⁰For 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.

8 Unreliable narration and imaginative resistance

(169) Maxine owns a donkey.

	x', y'	W
(170)	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.

	x', y'	W
(172)	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:

	x', y'	W
(173)	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')	
$1 > \{2, 3\} > \{4, 5, 6\}$	

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

8 Unreliable narration and imaginative resistance

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'} beat(x',y')	
	$\{0,1\} > \{2,3\} > 7 > \{4,5,6\}$	

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 speaker¹¹ 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

¹¹Note 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.

8.3 Fiction updates and belief revision

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)

	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'} beat(x', y')	
	$\{0, 1\} > \{2, 3\} > 7 > \{4, 5, 6\}$	

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: $K \uplus \phi = \text{merge } K \text{ with } \phi$, resolve all anaphora and presuppositions, and extend the epistemic entrenchment ranking to the new conditions.
 - b. contraction: $(K \uplus \phi) \div \perp = \text{remove as many low ranked conditions from } K \uplus \phi \text{ as needed to ensure that } \text{CG}_E(K \uplus \phi) \text{ is consistent (where } E \text{ denotes in every world the set of people engaged in the discourse in that world, and } \text{CG}_E K \text{ entails } \forall x \in E(\text{ACCEPT}_x K), \forall x \in E(\text{ACCEPT}_x(\forall x \in E(\text{ACCEPT}_x K))), \dots)$
 - c. caution: for any condition ψ added in the expansion phase but subsequently removed in the contraction phase, update with the corresponding hedged condition $\text{ACCEPT}_{s?}\psi$ (where ' $s?$ ' is an anaphor that needs to be bound to the current speaker), i.e., $((K \uplus \phi) \div \perp) + \text{ACCEPT}_{s?}\psi$

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

8 Unreliable narration and imaginative resistance

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.

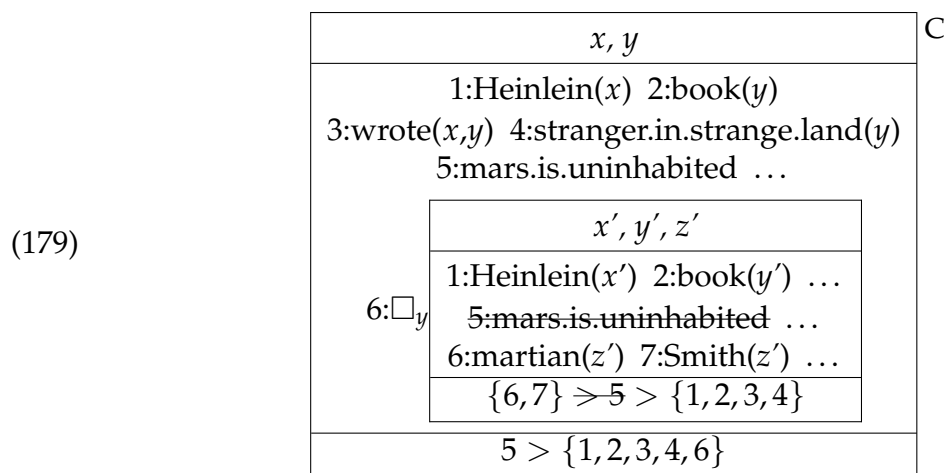
(178)

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')	
$\{6, 7\} > 5 > \{1, 2, 3, 4\}$	

8.4 Interpreting fiction

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:



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

¹²Perhaps joking is a form of narrative fiction, in which case we'd no longer be engaging in factual inquiry but fiction.

¹³Talking 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.

¹⁴Recall, Cautious Update is not triggered because we're retracting only old information (see (177)).

8 Unreliable narration and imaginative resistance

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

8.4 Interpreting fiction

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.

	x', y', u', s'	W
(182)	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.

8 Unreliable narration and imaginative resistance

(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'} grumbling.over.food(u')	
	$5 > \{6, 7, 8, 9, 10, 11, 12\} > \{1, 2, 3, 4\}$	

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

8 Unreliable narration and imaginative resistance

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')	
$\{5, 6, 7, 8, 9\} \geq 4 > \{1, 2, 3\}$	

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:¬	like.animals(u')	8:pour.bleach(u')
9:did.good(u')	10:ACCEPT _{s?}	did.good(u')
$\{5, 6, 7, 8, 10\} > 4 \succ 9 > \{1, 2, 3\}$		

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?

¹⁵For 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.

¹⁶We 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.

8 Unreliable narration and imaginative resistance

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)

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') ... 10:ACCEPT _{x'} did.good(u')	
$\{5 \dots 10\} > 4 > \{1, 2, 3\}$	

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)

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') ... 10:ACCEPT _{x'} did.good(u')	
$\{5 \dots 10\} \triangleright 4 > \{1, 2, 3\}$	

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 storyteller 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

¹⁷Though 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 need to establish a communion behind a narrator’s back.

Alternatively, the anaphoric subject of the hedge $ACCEPT_{s'}$ 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

	$x', y', u', v' s'$	W
(190)	1:author(x') 2:fish.tank(y') 3:wrote(x',y') 4:killing.animals.is.wrong ... 5:Sara(u') 6:father(v',u') ... 10: $ACCEPT_{s'}$ did.good(u')	
	11:narrator(s')	
	$\{5 \dots 11\} > 4 > \{1, 2, 3\}$	

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

¹⁸Altshuler 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.

8 *Unreliable narration and 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.