AUSTRALASIAN JOURNAL OF PHILOSOPHY, 2018
https://doi.org/10.1080/00048402.2018.1435698

OPEN ACCESS Check for updates

Truth in Fiction, Impossible Worlds, and Belief Revision

Christopher Badura[a] and Francesco Berto[b]

[a]RUB Research School & Department of Philosophy II, Ruhr-University Bochum; [b]Institute for Logic, Language and Computation, University of Amsterdam

ABSTRACT
We present a theory of truth in fiction that improves on Lewis's [1978] 'Analysis 2' in two ways. First, we expand Lewis's possible worlds apparatus by adding non-normal or impossible worlds. Second, we model truth in fiction as (make-believed) belief revision via ideas from dynamic epistemic logic. We explain the major objections raised against Lewis's original view and show that our theory overcomes them.

ARTICLE HISTORY Received 17 June 2017; Revised 10 January 2018

KEYWORDS truth in fiction; impossible worlds; inconsistent fiction; Sylvan's box, belief revision

1. Introduction

Fictions are generally not true, but we can truthfully talk of what happens in them—including what they are not explicit about. When Heathcliff and Catherine meet for the final time in *Wuthering Heights*, Heathcliff is dressed in the manner of an eighteenth-century country gentleman, not as a circus clown. Yet the text of that scene says nothing about how Heathcliff is dressed. So, what is true in the fiction of *Wuthering Heights* goes beyond what's explicitly written in the text. Such *going beyond* is difficult to model.

The modal analyses proposed by Lewis [1978] have been the go-to approach for truth in fiction. However, they have faced serious objections (for a summary, see Sainsbury [2010: ch. 4]). We improve on Lewis's 'Analysis 2' in two ways. First, we expand his possible worlds apparatus by adding non-normal or impossible worlds that represent some impossibility as being the case [Kiourti 2010; Berto 2012; Nolan 2013; Jago 2014]. Second, we take ideas from belief revision theory [Grove 1988; Segerberg 2001; Baltag and Smets 2011] to give truth conditions for sentences of the form ⌜In fiction *f*, *A*⌝ that avoid all of the main criticisms faced by Lewis.

In section 2, we introduce Lewis's Analyses 0, 1, and 2, and some major objections advanced against them. In section 3, we deal with a difficulty that depends on Lewis's working only with possible worlds. In section 4, we present our improved version of Analysis 2. In section 5, we show how it solves Lewis's problems.

2. Lewis on Truth in Fiction

We say 'In *Star Trek*'s world…' and 'In the world of *The Lord of the Rings*…'. We may thus understand truth in fiction along the lines of truth relative to a world. One may take the world of *Wuthering Heights* as one that is compatible with everything stated explicitly in the novel. But unreliable narrators can explicitly state something that turns out, later on, not to be true in the fiction, or can make claims tongue-in-cheek. We shall only take into account reliable narration, in which a narrator's explicitly claiming that *A* is sufficient for *A*'s being true in the fiction. This seems the default situation when we engage with works of fiction.[1] We thus endorse this principle as a default rule:

> EXPLICIT. If $\ulcorner A \urcorner$ occurs explicitly in the story *f*, then $\ulcorner A \urcorner$ is true in *f* and so also $\ulcorner In_f, A \urcorner$ is true.[2]

'In_f, A' abbreviates 'In the fiction *f*, *A*', where '*f*' is a placeholder for the title of some fiction[3] and 'A' is a placeholder for (depending on context) a sentence or a proposition.[4]

The worlds complying with the explicit content of *Wuthering Heights* are all worlds in which Heathcliff is adopted by Mr Earnshaw, comes to own Thrushcross Grange, marries Isabella. We speak of world*s* now, as Lewis [1978] proposes that we associate a plurality of them with the fiction: although the explicit content of the novel narrows down the class of worlds to those compatible with it, we cannot fulfil our initial hope to get *the* unique world of *Wuthering Heights*. Given fiction *f*, there generally are sentences *A* such that neither is it true in *f* that *A* nor is it true in *f* that $\neg A$. We cannot distinguish between a world where Frodo has an even number of hair on his feet and a world where that number is odd. In a formal representation, that is $\neg In_f, A \wedge \neg In_f, \neg A$. So, for some fiction *f* and sentence *A*, we have $\neg(In_f, A \vee In_f, \neg A)$.[5]

Lewis [ibid.: 39] also claims that we should individuate fictional stories by looking at particular acts of storytelling: fictions with the same explicit content will differ when uttered by different authors on different occasions. This suggests that acts of storytelling are individuated by their respective utterers as well as by the place and time of utterance. This, however, generates a tension with Lewis's proposal, one page later, on how to select the relevant worlds [ibid.: 40]:

> The worlds we should consider, I suggest, are the worlds where the fiction is told, but as known fact rather than fiction. The act of storytelling occurs, just as it does here at our world; but there it *is* what here it falsely purports to be: truth-telling.

Following Searle [1975], for Lewis there is a distinction between what the act of storytelling amounts to in our world and what it amounts to in other worlds. At our world, the author is pretending to tell something as known fact, or to be someone telling the story as known fact. At other worlds, someone is uttering something lexically identical but telling it as known fact. In writing *Wuthering Heights*, Brontë is not performing an illocutionary act of assertion. She is pretending to be someone, Lockwood or Nelly, who is asserting the story.

[1] Heyd [2006] proposes to account for unreliable narration pragmatically via Gricean maxims. Unreliable narration is much discussed in literary theory (see Riggan [1981] and Köppe and Kindt [2011]).

[2] We will normally avoid corner quotes when context disambiguates.

[3] Under which conditions a work counts as a fiction, or two fictions are identical, is often not so clear. For discussion of the former, see Searle [1975], Currie [1990], Friend [2012], and Matravers [2014]. For the latter, see Löwe [2010]. We lack the space here to deal with these issues.

[4] Following Folde [2011], '$\ulcorner In_f, A \urcorner$ is true', 'it's true in fiction *f* that *A*', and 'in the fiction *f*, it is true that *A*' will be treated as equivalent.

[5] Incompleteness with respect to *A* does not entail that, for some *f* and *A*, we have $\neg(In_f, A \vee \neg In_f, A)$. But if a fiction *f* is incomplete with respect to *A*, then $\neg In_f, A$; thus, given $\vee$-introduction, $In_f, A \vee \neg In_f, A$ for the particular *A* and *f*. Also, *f*'s incompleteness with respect to *A* should not entail that $In_f, \neg(A \vee \neg A)$. Although neither In_{LOTR} (*The Lord of the Rings*), Frodo is right-handed, nor In_{LOTR}, Frodo is left-handed, it seems true that In_{LOTR}, (Frodo is left-handed $\vee$ Frodo is right-handed). This does not rule out the possibility of fictions *f* and sentences *A*, such that, $In_f, \neg(\mathrm{A} \vee \neg A)$.

Then the sameness of the acts cannot be based on their being the same kind of speech act. Moreover, the story is told by different people: at our world by the author, and at other possible worlds by the narrator, who is not necessarily (cross-worldly) identical to the author. So, we should have different acts of storytelling merely because they correspond to different acts of utterance. Thus, it seems to us, the sameness of the act(s) of storytelling can only lie in their lexical identity: we should consider worlds *w* such that if in the actual world the author of *f* pretends to tell *f* as known fact in act *a*, then in *w* the narrator of *f* tells *f* as known fact in act a', where *a* and a' are lexically identical. This excludes the actual world as a candidate for one of our chosen worlds. For even when author and narrator are identical, pretending to assert *A* and asserting *A* cannot occur in the same act of utterance [ibid.].

Now for the core issue: the implicit content of a fiction. Lewis starts with Analysis 0: $\ulcorner \mathrm{In}_f, A \urcorner$ is true iff at every world where the story *f* is told as known fact, rather than as fiction, *A* is true. Whenever the author pretended to assert that *A*, the narrator, in telling *A* as known fact, asserts *A*. Together with our restriction to reliable narrators, this implies that the worlds where the story is told as known fact are those where the explicit content is true.

Because in Analysis 0 the fictional operator is just a restricted quantifier over possible worlds, whatever necessarily follows from a truth in the fiction is itself true in the fiction. It will be true in *Wuthering Heights* that (assuming classical logic) either Heathcliff loves Catherine or doesn't, and (assuming that the mental metaphysically depends on the physical) that his tortured mental life supervenes on various physical facts about him. Any necessary truth will be true in any fiction: that Fermat's Last Theorem holds, that there exist properties, that truth is grounded in obtaining states of affairs (assuming that these are metaphysical truths). Proudfoot [2006] takes these as unwelcome results; we will return to them.

There must be further constraints on the set of worlds of *Wuthering Heights*. It might seem intuitive that gravity obeys an inverse square law in *Wuthering Heights*, just as it does in reality, although Brontë never says so explicitly. We might take the worlds of *Wuthering Heights* to be those nomologically possible worlds compatible with what's explicitly stated in the text. Also, to grasp the novel we need to understand something about the social customs of late eighteenth-century England. Heathcliff leaves, becomes wealthy, returns as a gentleman. To get the importance of those events, we need to understand something of England's attitudes to class and gender in the eighteenth century. Brontë took for granted that her contemporary readers would easily get these. So, we may further restrict the worlds of the fiction to those compatible with various social norms.

If we restrict to nomologically possible worlds, though, how are we to understand Catherine's ghost? Ghosts are physically impossible. On this way of constraining, we must say that the narrator (Lockwood) is either lying (hence unreliable) or hallucinating. This may fail to make good sense of *Wuthering Heights*. The approach makes even worse sense of any ghost story with a reliable narrator (for example, *Canterville Ghost*), where nomologically impossible events are central to the plot.

As for social customs, Will Self's *Great Apes* tells the story of a man who wakes up as an ape, in a society in which apes are socially superior to humans. Many of our social norms are turned on their head: a good Catholic should be as sexually promiscuous as possible. The worlds of *Great Apes* don't share our social rules.

Overall, on the one hand, Analysis 0 doesn't import enough: it does not account for some relevant implicit content—namely, background beliefs or knowledge we import into the story, unless those logically follow from the explicit content. On the other hand, it imports too much—namely, any logical and metaphysical truth.

Lewis's Analysis 1 piggy-backs on the Lewis-Stalnaker semantics for counterfactuals: it takes the worlds of the fiction as those that are closest, or most similar, to our own reality, whilst respecting what's stated explicitly in the text. Call the worlds where *f* is told as known fact, '*f*-worlds'; those where some sentence *A* is true, '*A*-worlds'; those where the explicit content obtains and *A* holds, '*fA*-worlds'. Analysis 1 goes thus: $\ulcorner \mathit{In}_f, A \urcorner$ is true iff there is an *fA*-world that is, on balance, more similar to our world than any *f*-world that is not an *A*-world.

Because we have only possible worlds, we still import every logical and metaphysical truth into any fiction. Also, where Analysis 0 did not allow for enough imports, Analysis 1 seems to allow for too many (see, for example, Currie [1990] and Proudfoot [2006]). Where '*SH*' stands for the Sherlock Holmes stories, consider this:

- In_{SH}, Trump wins the presidential election in 2016.

By Analysis 1, this is true iff some *SH*-world *w* in which Trump wins the election in 2016 is closer to our world than is any *SH*-world where Trump does not. With there being such a *w*, one is saying that *SH*-stories are, and were at the time of writing, about Trump and our future. But how can the *SH*-stories be about (say) flying cars in 2160? Call this the *aboutness-objection*.

Lewis's Analysis 2 restricts the worlds at which we look, not to those objectively most similar to the actual one, but to the subjectively most plausible given the *overt beliefs* of storytellers and their contemporary audience (the 'community of origin' of the fiction). Overt beliefs are what almost everyone believes, what almost everyone believes that almost everyone believes, and so on. Some false things may count as true in a story because they were among the overt but false beliefs of the author and the audience. Take some mediaeval tale never stating explicitly that the Sun revolves around the Earth: one may still want to count this as true in the tale, because it was an overt belief at the time.

Call the worlds where all of the overt beliefs of the community of origin of the story are realized, 'overt-belief worlds'. Analysis 2 goes thus: $\ulcorner \mathit{In}_f, A \urcorner$ is true iff, for every overt-belief world *w*, there is an *fA*-world that is closer to *w* than is any *f*-world that is not an *A*-world. As there are overt-belief worlds where Trump wins the election and overt-belief worlds where he doesn't, it is neither the case that in the *SH* stories he wins the election, nor the case that in the *SH* stories he doesn't. Although this helps with the Trump problem, it's still the case that in the *SH* stories either Trump is elected as president or he isn't. The aboutness-objection strikes again.

Analysis 2 may also import actually false beliefs into stories in which these beliefs are false as well [Bonomi and Zucchi 2003]. Take a novel written in Nazi Germany by a progressive author, Kurt, opposing the regime. Suppose that the main character, Shlomi, is Jewish, while the antagonist, Becker, is a Nazi. Analysis 2 makes true in the story that Shlomi is a lesser human being than Becker. This doesn't seem to be true in the story, given Kurt's convictions.

The aboutness-objection and the import of logical-metaphysical truths are difficulties affecting all of Analyses 0, 1, and 2: they are spin-offs of Lewis's having only possible worlds in his analyses. In the next section, we reach the main trouble with this set-up.

3. Inconsistent Fictions

Each Lewisian analysis requires possible worlds where the story is told as known fact. But it doesn't seem to be the case that we can tell something as known fact if it is impossible: knowledge implies truth. If there are no possible worlds where *f* is told as known fact, then $\ulcorner In_f, A\urcorner$ is vacuously true for any *A*. Watson's war wound in some Holmes stories is on his shoulder; in others, it's in his knee. It's also a given that he only has one wound. One wound that is two wounds? Then everything comes out true in the Holmes stories.

Lewis [1983] proposes two solutions—the method of union and the method of intersection. We can look at what's true (according to Analysis 1 or 2) in some, or in all, of the maximally consistent fragments of an inconsistent story. On the former method, it's true in the Holmes stories that Watson's only war wound is on his shoulder and it's true in the stories that his only war wound is on his knee, but the conjunction of the two is not true in the stories. On the latter method, neither is true.

But some inconsistencies in fiction are no narrative oversight [Priest 1997: 575–6]:

> Carefully, I broke the tape and removed the lid. The sunlight streamed through the window into the box, illuminating its contents, or lack of them. For some moments I could do nothing but gaze, mouth agape. At first, I thought that it must be a trick of the light, but more careful inspection certified that it was no illusion. The box was absolutely empty, but also had something in it. Fixed to its base was a small figurine, carved of wood, Chinese influence, Southeast Asian maybe.
>
> I put the lid back on the box and sat down hard on the armchair, my mental states in some disarray. I focused on the room. It appeared normal. My senses seemed to be functioning properly. I focused on myself. I appeared normal. No signs of incipient insanity. Maybe, I thought, it was some Asian conjuring trick. Gently, I reopened the box and gazed inside. … The box was really empty and occupied at the same time. The sense of touch confirmed this.

In Graham Priest's *Sylvan's Box*, the narrator is Priest himself, or a fictional version of him. As author, he asks us: what's true in this fiction? Taking the narrator's explicit statements at face value, it is true in *Sylvan's Box* that Priest discovers a box that is empty but has something in it. The obtaining of a contradiction is essential for understanding the story. But neither the method of intersection nor that of union deliver that it's true in *Sylvan's Box* that a box is *fempty* (both full and empty).

Hanley [2004] and Nolan [2007] have challenged the view that a contradiction's being true in *Sylvan's Box* is essential to the plot. Both opt for a reading under which the narrator, Priest, falsely believes that there is a fempty box. But, first, such an interpretation doesn't help to make sense of Priest's description of how everything appears normal when not looking into the box, the fact that his sense of touch confirms his belief, and that his colleague confirms Priest's perception. The interpretation is a result of considerations on how to consistently make sense of what is explicitly stated. It is some inference to the best explanation—where 'best', for Hanley and Nolan, means or entails 'as close as possible to the real world'. This position seems to be motivated by the fact that, up until Priest opens the box, the story is perfectly realistic.

But so are many of Stephen King's stories. Still, not each of his works supports the hypothesis that the narrator is unreliable, the protagonist is merely going insane, or is in a nightmare. Sometimes, the best interpretation is that a realistic story-line suddenly turns into something sharply diverging from our reality. Why not so in Priest's case? Many theories of interpretation put emphasis on the text and what's written down explicitly. So, what's the *best* inference is what stays as close as possible to the explicit text. And an inference to that kind of best explanation does not support the hypothesis of Priest having a false belief or being an unreliable narrator.

Second, even if the interpretation worked for *Sylvan's* Box, there seem to be other cases—for example, the movie *Last Action Hero* (*LAH*) [Proudfoot 2006]—where a character that in *LAH* is fictional, in *LAH* becomes human. So, in *LAH*, he's human and fictional—a metaphysical impossibility. Claiming that Danny, the protagonist, is hallucinating in the story, or that the depiction on the screen is meant to be deceptive, doesn't make sense of the plot.

Given Explicit and the truth of a contradiction in *Sylvan's Box*, it is still not true in the story that Priest finds himself levitating, dressed in a tutu. Yet that would be classically entailed by the box's being fempty. So, what's true in this fiction isn't closed under classical entailment. Priest has it that 'the logic of the story must be paraconsistent' [Priest 1997: 580]. It seems that *some* inferential principles must apply. In the story, the box is empty. And, in the story, the box has something in it. Those truths of the fiction are, strictly speaking, not explicit in the text. What's explicit is their conjunction (expressed with 'but'). Yet it seems clear that the conjuncts are also true in the fiction. Further exploration of the text may provide evidence that the standard introduction and elimination rules for conjunction, disjunction, and the conditional, plus double negation introduction and elimination, are all fine in *Sylvan's Box*. This might give reason to think that the logic of this fiction is some paraconsistent logic—for instance, $\mathrm{LP}^{\rightarrow}$ by Girard and Tanaka [2016], to which we shall return.

However, we have to distinguish between the question of what the logic is of a particular work of fiction, such as *Sylvan's Box*, and the question of what the general logic is of fictional operators. It is the latter, if there is such a thing, that deserves the label 'the logic of fiction'. The former is, rather, the logic of *a* fiction. Now, we are not sure that *any* logic may be singled out as *the* logic of fiction: for any fiction obeying logic *L*, we may be able to write a logical fiction where some principle or other of *L* fails. Our modelling, to which we come in the next section, should allow for the possibility of any logic (save perhaps for the trivial logic whose only validity is the entailment from *A* to *A*) failing in some fiction.

4. A Formal Semantics

Our account is based on Lewis's Analysis 2, expanded via the addition of (a) impossible worlds and (b) an apparatus modelling our understanding of fiction as a kind of *belief revision*. Let us have a domain of possible and impossible worlds, totally ordered by a plausibility relation, as in the semantics for doxastic and epistemic logics of belief revision (see, for example, Grove [1988] and Segerberg [1995, 2001]). Think of the ordering in terms of nested spheres around a core, as in the standard Lewis semantics for counterfactuals, except that the spheres don't model objective similarity. They model subjective plausibility, or degrees of belief entrenchment. The innermost sphere is the set of worlds that realize certain beliefs. The closer the other worlds are to the core, the more plausible they are for the relevant agent(s): they are more likely to be embraced as fallback positions after belief revision induced by new information.

To the best of our knowledge, the application of impossible world semantics to truth in fiction has never been worked out in formal detail, the closest being Priest [1997]. The only modellings of which we are aware that use ideas from formal belief revision theory are Rapaport and Shapiro [1995] and Klassen et al. [2017]. Our work differs from both of these in the use of soft upgrades for belief revision, the consideration of common belief, and the use of impossible worlds.[6]

Nichols and Stich [2003] give a cognitive model of pretence that resembles ideas from belief revision theory. Acts of pretence in imagination have a deliberate starting point: 'an initial premiss or set of premises, which are the basic assumptions about what is to be pretended' [ibid.: 24]. But also, 'children and adults elaborate the pretend scenarios in ways that are not inferential at all', integrating the explicit content with 'an increasingly detailed description of what the world would be like if the initiating representation were true' [ibid.: 26–8]. Our cognitive architecture, according to Nichols and Stich, comprises an *(im)possible world box* accessed when we engage in the pretence via the pretence premises, and a belief box from which we select background beliefs to import in order to integrate the pretended scenario. We then consider those worlds where the pretence premises plus the imported beliefs obtain (and are closed under some logic). However, according to Nichols and Stich, we quarantine the pretence premises by indexing them, so that we know how to reroll the update after the pretence. In our formal modelling, we omit the indexing for simplicity. The overall content of the pretence is what holds in the worlds where the explicit pretence premises are true, *and* which are the most plausible in terms of our background beliefs, adjusted to make room for the explicit pretence premise itself.

Now, this looks a lot like the *soft upgrades* of doxastic-epistemic logics of belief revision [van Benthem 2007; Baltag and Smets 2011]:[7] agents order the set of worlds, based on how strongly they take the various worlds as candidates for actuality compared to the state they are in (see van Ditmarsch [2005]).

Let '${}_w\leq$' stand for such ordering, and read '$v\ {}_w\leq u$' as 'v is at least as plausible with respect to w as u' (for a given agent). One typical kind of soft upgrade has it that incoming new information A gives a new ordering ${}_w\leq^A$ such that, for all worlds v and u, if A is true in v and A is not true in u, then $v\ {}_w\leq^A u$. Let W be the set of all worlds. If, for every $u \in S \subseteq W$, we have that $v\ {}_w\leq u$ (or $v\ {}_w\leq^A u$), we say that v is ${}_w\leq$-minimal (${}_w\leq^A$-minimal) in S, or simply minimal in S if the ordering and the subset are clear from context.

Now just think of an agent reading a fiction as facing new information with every sentence it is reading, thus sequentially upgrading its beliefs with each sentence of the fiction. The agent need not actually believe what it is reading. One can easily incorporate the aforementioned quarantining suggested by Nichols and Stich by indexing the worlds by their previous position in the ordering. When the agent stops engaging with the fiction and is checked for its actual beliefs, it will recover the old ordering based on the index.[8]

[6] Thanks to Toryn Klassen for pointing out these references.

[7] Girard and Tanaka [2016] provide a general approach for belief revision in any many-valued logic, with an emphasis on LP. Hence, they model belief revision with a non-normal base logic. We return to this below.

[8] Alternatively, one can index the sentences with which it upgraded and, when rerolling the upgrade, it will upgrade with the negation of the indexed sentences. In this way, one could also incorporate ideas from Searle [1975] and Currie[1990], where the reader recognizes the intention of the author as to whether a sentence is supposed to be make-believed or believed. If the agent is supposed to believe A, it might upgrade not with the indexed A' but rather with A itself and, when rerolling the upgrade, A-worlds that are also candidates for the real world will be more plausible than before reading the fiction. Thus, the agent has learned something new through the fiction.

Lewis's Analysis 2 refers to the overt beliefs of the community of origin. We change this in two ways. First, rather than overt beliefs we take *common* ones: that is, the beliefs such that everyone has them, everyone believes that everyone has them, and so on. This is a merely practical choice: it avoids the 'most' quantifier coming with overt beliefs for a plain 'all' in our analysis below. Second, we take the community of *interpretation* of a work of fiction, instead of that of origin. We want to model this community's reasoning on the fiction. Such a shift of focus will turn out to be useful to address a dilemma that we will meet in the following section. Overall, when we evaluate ⌜In_f, A⌝ for truth we look at all those worlds, possible or not, that are minimal with respect to the common belief worlds of the community of interpretation after the upgrade with the explicit content of *f*.

Take a propositional language L with the usual set of atoms AT, negation $\neg$, conjunction $\wedge$, disjunction $\vee$, a strict conditional $\prec$, modal operators $\Diamond$ and $\Box$, and a family of 'in fiction *f*'-operators (one for each *f*), In_f. The well-formed formulas are the atoms plus, if A and B are formulas:

$$\neg \mathrm{A} \mid (A \wedge B) \mid (A \vee B) \mid (A \prec B) \mid \Box B \mid \Diamond \mathrm{B} \mid In_f,\ A$$

The explicit content F of fiction *f* is a *finite* sequence of sentences that express what is expressed by telling *f* as known fact rather than as fiction. Thus, it goes beyond the literal meaning of *f* via a treatment of metaphor, irony, etc., that we suppose is given by pragmatics. We also write, abusing notation, $\wedge_{A\in F}A$ for F. Thus, we treat the explicit content as a finite conjunction.

A *multi-agent plausibility frame* is a tuple $\mathcal{F} = <W, N, \mathrm{Ag}, \{{}_{w}\leq_{a} | w\in W, a\in \mathrm{Ag}\}>$. W is a set of worlds, N is the set of normal-possible worlds, W–N is the set of non-normal worlds, Ag is a finite set of agents, each ${}_{w}\leq_{a}$ is agent a's plausibility ordering on W in state w. We assume each ${}_{w}\leq_{a}$ to be transitive and well-founded: for every set of worlds $S\subseteq W$, we can always determine the most plausible worlds in S: ${}_{w}\leq_{a}$ -min(S) = $\{v\in S| \forall x\in S: v\ {}_{w}\leq_{a}\ x\} \neq \{\}$. Well-foundedness of ${}_{w}\leq_{a}$ guarantees that it is a reflexive and total order on W. The latter means that any two worlds are comparable. We define $v\ {}_{w}<_{a}\ u$ iff. $v\ {}_{w}\leq_{a}\ u$ and not $u\ {}_{w}\leq_{a}\ v$. We define $v\ {}_{w}\cong_{a}\ u$ iff. $v\ {}_{w}\leq_{a}\ u$ and $u\ {}_{w}\leq_{a} v$.

A frame becomes a model $\mathcal{M} = <W, N, \mathrm{Ag}, \{{}_{w}\leq_{a}| w\in W, a\in \mathrm{Ag}\}, \rho>$ when endowed with an evaluation relation ρ, relating at worlds the atoms in AT to truth ('$\rho_w p1$'), falsity, ('$\rho_w p0$ '), both, or neither – this mirrors the 'relational semantics' for the logic of First Degree Entailment, and its expansions (see Priest [2008]).

Before we extend ρ to the whole language, let us justify why we use an FDE-based framework. The previous section motivated that ρ can relate worlds to both truth values. That ρ can relate a world to neither truth value is based on the following ideas. First, if there are stories that make contradictions true because we adhere to EXPLICIT, we do not see an immediate reason to deny that one could write a story *f* s.t. In_f, $\neg(A\vee\neg A)$ by explicitly writing such a story. Second, given the aboutness objection, we might say that every fiction remains incomplete wrt some sentences; not only in the sense of $\neg In_f, A \wedge\neg In_f\neg A$ (cf. note 5) but in the sense that all *f*-worlds are incomplete worlds wrt A—they do not decide A because the story isn't about A at all, and so aren't any of the story-worlds. Finally, Bourne and Caddick [2016] have argued that time in fiction requires the indeterminate truth value. Our framework provides the option of modelling all these.

If one is not convinced by these arguments, one could rely on a dynamics on top of a static LP-base, as developed by Girard and Tanaka [2016]. They provide a general setting suitable for any many-valued logic. So, our approach could be seen as a special case. However, as we explain below, we define beliefs, common beliefs, and the upgrade for the meta-language to account for the truth-conditions of In_{f}, A, whereas they provide a semantics for belief and upgrade in the object-language. Nevertheless, their approach seems to be applicable to the issue of truth in fiction, using our idea of modelling truth in fiction as soft upgrades on a non-classical static base.

We extend ρ to the whole language as follows. For the extensional connectives we have, for all $w \in N$,

(S1¬) $\rho_w(\neg A)1$ iff. $\rho_w\, A0$
(S2¬) $\rho_w(\neg A)0$ iff. $\rho_w A1$
(S1∧) $\rho_w(A \wedge B)1$ iff. $\rho_w A1$ and $\rho_w B1$
(S2∧) $\rho_w(A \wedge B)0$ iff. $\rho_w A0$ or $\rho_w B0$
(S1∨) $\rho_w(A \vee B)1$ iff. $\rho_w A1$ or $\rho_w B1$
(S2∨) $\rho_w(A \vee B)0$ iff. $\rho_w A0$ and $\rho_w B0$

The familiar modalities get their usual (S5) clauses, over normal worlds. For all $w \in N$,

(S1≺) $\rho_w(A \prec B)1$ iff. for all $v \in N$, if $\rho_v A1$, then $\rho_v B1$
(S2≺) $\rho_w(A \prec B)0$ iff. for some $v \in N$, $\rho_v A1$, and $\rho_v B0$
(S1□) $\rho_w \Box A1$ iff. for all $v \in N$, $\rho_v A1$
(S2□) $\rho_w \Box A0$ iff. for some $v \in N$, $\rho_v A0$
(S1◇) $\rho_w \Diamond A1$ iff. for some $v \in N$, $\rho_v A1$
(S2◇) $\rho_w \Diamond$ 0 iff. for all $v \in N$, $\rho_v A0$

We say that A is a logical consequence of a set of formulas Γ ($\Gamma \Vdash A$) iff, for every $w \in N$ in every model $\mathcal{M}$, whenever, for all $B \in \Gamma$, $\rho_w B1$, then $\rho_w A1$. A is a logical truth iff $\{\} \Vdash A$.

We impose a (so-called) *Classicality Condition*: for every $w \in N$ and every atom p, we have ρ_wp1 or ρ_wp0 but not both. This extends to every formula by induction. As for worlds in W–N, at them ρ relates logically complex formulas to truth values *directly*, not recursively (this is a common move in impossible worlds semantics: see, again, Priest [2008]). Worlds in W–N are not subject to the Classicality Condition.

The following version of Nolan's principle [Nolan 1997] allows us to have enough worlds at our disposal in the models:

NP. For any two formulas A, B, there are worlds w, $v \in W$, such that $\rho_w A1$ and not $\rho_v B1$.

We define the notion of common belief in our *meta*language as follows. Agent a believes a formula A at $w \in N$, $\mathrm{Bel}_a\, A$ (at $w \in N$), if A is true in every world in ${}_w\!\leq_a$-min (W), the worlds that a considers most plausible in the model from state $w \in N$. Everyone in a group G of agents believes a formula A (at a w∈N), $E_G A$, iff $\wedge_{a \in g}\, \mathrm{Bel}_a$ A is true at w: that is, everyone in fact believes A at w. Let us abbreviate $E_G E_G{}^{n\text{-}1} A$ with $E^n{}_G A$, and define $E^0{}_G A = A$. Then A is commonly believed in a group G (at $w \in N$), $\mathrm{CBel}_G A$, iff $\wedge_{i=1}{}^{\infty} E^i A$ is true at w in the model: that is, (at w) everyone believes A, and everyone believes that everyone believes A, and so on.

A world $v \in W$ is a common belief world of a group G at world w if everything that's commonly believed at w is true at v. So, v is a common belief world of group G at world w if, for every A such that at w it is true that $\mathrm{CBel}_G A$, A is true at v. Let us denote the set of common belief worlds at w by $|\mathrm{CBel}_G{}^w|$.

A soft upgrade with A reorders the worlds in such a way that all of those in which A is true are then considered more plausible than all of those where A is not true. We obtain the new ordering as follows, where A is from F of f:

(C1) For all u, $v \in W$, if $\rho_u A1$ and it is not the case that $\rho_v A1$, then $u_w \leq_a{}^A v$.
(C2) Otherwise, the old ordering remains.

Given NP, we have some world in the model in which A is true. So, the upgraded relation will be non-empty. Also, it preserves transitivity and well-foundedness.

We can add additional constraints on ${}_w\leq_a{}^A$: for example, that (non-normal) worlds that make A true *and* false are less plausible than those making A only true. Or, if the agent follows some pragmatic rules of interpretation, certain worlds obeying those rules might be considered more plausible (for such a pragmatic-based approach, see Bonomi and Zucchi [2003]). But for simplicity, let's just stay with the above conditions.

We are interested in a multi-agent setting, but, given the individual orderings, it is no trivial task to come up with a group ordering $\leq_G$. This should reflect how the agents can agree, for every set of worlds, on some set of most plausible or preferred worlds. This amounts to a voting problem among infinitely many, or at least arbitrarily finitely many, alternatives. One has to be aware here of the impossibility result for social choice functions shown by Arrow [1950] and the analogue for belief revision shown by Leitgeb and Segerberg [2007]. Both assume certain plausible conditions on the group of agents, and then show that a group preference ordering cannot be obtained. One of these conditions is non-dictatorship, which, on some theory of interpretation, might be relaxed to give some priority to the author's ordering. This is the way we go here, assuming a hierarchy among the agents a_0, ..., a_n where a lower index indicates a higher rank.[9] Let Gi be the group of agents from a_0 up to and including agent a_i and F as the explicit content. Then we generalize ${}_w\leq_G{}^F$, the group ordering after revising with the explicit content of the fiction, as follows:

$$
\begin{aligned}
&{}_w\leq_{G0}{}^F = {}_w\leq_{a0}{}^F\\
&{}_w\leq_{Gn+1}{}^F = {}_w<_{Gn}{}^F \cup \left({}_w\cong_{Gn}{}^F \cap {}_w\leq_{an+1}{}^F\right)
\end{aligned}
$$

If, for every agent a∈Gi, the individual orders are transitive and well-founded and satisfy C1 and C2, then it can be shown that ${}_w\leq_G{}^F$ is transitive, well-founded, and satisfies C1and C2. (As an alternative to giving up non-dictatorship, one can plug in one's

[9] The interpretation of the ranking is actually difficult, as ranking by competence leads to taking into account a less competent agent when the more competent one is indecisive. However, one might say that it is a ranking based on authority, where the author comes first and then the respective literary experts follow. This presumes a particular theory of interpretation, giving a special role to the author, which can indeed be challenged (see, e.g., Barthes [1977] and Gendler [2011]). We lack the space to address the challenge in this paper.

favourite method of giving up conditions on the orderings to obtain a group ordering and to avoid the impossibility results.)

With this group ordering, we can, for every $S \subseteq W$ and A from F, have this:

$$_{\mathrm{w}}{\leq_{\mathrm{G}}}^{\mathrm{A}}\text{-min}(\mathrm{S}) = \left\{v \in S \mid \forall \mathrm{x} \in S : v_{w}{\leq_{\mathrm{G}}}^{\mathrm{A}}\mathrm{x}\right\}$$

This is the set of worlds considered most plausible by the group with respect to the set S after upgrading with A.

Here comes, finally, our definition of truth in fiction. For $w \in N$,

$$\begin{array}{ll}(\mathrm{S1}\, In_f) & \rho_{\mathrm{w}}(In_f,\ A)1 \text{ iff, for every world } v \in {}_{\mathrm{w}}{\leq_{\mathrm{G}}}^{\mathrm{A}}\text{-min } (|\,\mathrm{CBel_G}{}^{\mathrm{w}}\,|),\ (\rho_{\mathrm{v}}\mathrm{F}1 \Rightarrow \rho_{\mathrm{v}}A1) \\ (\mathrm{S2}\, In_f) & \rho_{\mathrm{w}}(In_f,\ A)0 \text{ iff not } \rho_{\mathrm{w}}(In_f,\ A)1\end{array}$$

By '$\rho_{\mathrm{v}}\mathrm{F}1$ ', we mean to express that the explicit content of f is true at v (*qua* being told as known fact at v). Thus, A is true in fiction f iff every f-world, considered by group G most plausible with respect to the common beliefs of G after revising with the explicit content of the fiction, makes A true (and dually for the falsity condition). We have In_f, $\wedge_{\mathrm{A}\in\mathrm{F}}\mathrm{A}$ (In_f, F) as a validity. The proof is obvious. If, in f, conjunction elimination holds, then, for every $A \in \mathrm{F}$, In_f, A is also a validity. Again, the proof is obvious.

As a consequence, to deny that it's true in *Sylvan's Box* that there is a box that is empty and not empty, In_{SB}, $(A \wedge \neg A)$, one has to argue for why the explicit occurrence of ⌜$A \wedge \neg A$⌝ is not sufficient for In_{SB}, $(A \wedge \neg A)$. Hence, truth in fiction becomes, at least in this case, dependent on theories of interpretation. We consider this a result that adequately models our reasoning about truth in fiction.

However, the logic will in general be very weak, for we have impossible worlds around. Counterexamples are easily given by choosing a(n impossible) world among the best worlds after the upgrade, and setting up the ordering and valuation just right. As we mentioned before, one can put further constraints on the upgraded plausibility ordering, which would likely generate more validities.

Having impossible worlds among the best ones helps a lot. Engaging with *Sylvan's Box* (*SB*) does not require us to revise our ordering much—up to the point where the fempty box appears. We say that the most plausible worlds are those where Priest has the right belief, and there actually is a fempty box, without trivialization. Easily, the inference from In_f, $(A \wedge \neg A)$ to In_f, B does not go through, specifically when the contradiction is part of the explicit content as per typical blatantly inconsistent fictions.

Here is a formal counterexample. For simplicity, we consider only two agents: Priest, p, and Nolan, n. So, G = {p, n}. Let the hierarchy between agents be p, n, for p is the author of *SB*. Let A be 'the box is empty.' The arrowhead points to the world considered at least as plausible wrt w. We omit reflexive arrows. Let w be the actual world. Here is the initial (toy)[10] model for *SB*:

Everyone considers the actual world most plausible. Nolan is indifferent between worlds that both make a contradiction true.

[10] There are going to be further worlds in the countermodels, in order for them to comply with NP. We don't specify them, for they don't do work, as far as exhibiting invalidities is concerned.

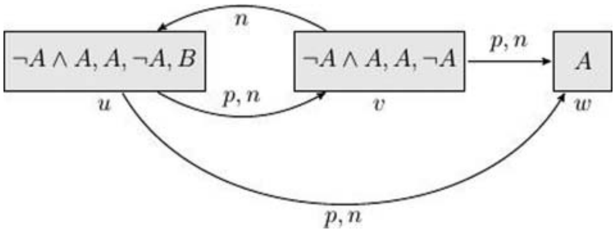

Figure 1.

At *w*, the agents believe *A*, and also at *w* it is commonly believed that *A*. The worlds *v* and *u* are also common belief worlds. After upgrading with *A*∧¬*A* we get the following model:

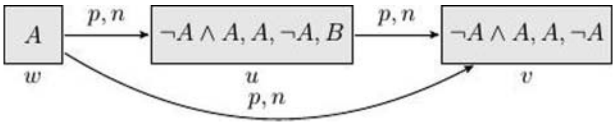

Figure 2.

As p was not indifferent between any worlds, the group ordering collapses into its individual upgraded ordering. Hence, we give maximal priority to the author's intentions. The most plausible common belief world that makes *A*∧¬*A* true is *v*, and so In_f, (*A*∧¬*A*) holds at *w*. But, at *v*, *B* does not hold. Hence, In_f, *B* is not true at *w*.

If we changed the hierarchy among p and n to n, p, the upgraded model would in fact remain the same, for the indifference of n between *u* and *v* is resolved by p's upgraded ordering.

The *Last Action Hero* case is modelled in a similar way. The most plausible worlds are going to be inconsistent but non-explosive. Moreover, non-normal or impossible worlds in our setting allow for *any* (non-trivial) logic to fail, and so not all tautologies will come out true in every story, taking into account Proudfoot's [2006] worry.

What about accidental contradictions? We treat the explicit content as one big conjunction. So, if both conjuncts explicitly occur in the explicit content, we most likely need to eliminate several other conjuncts and maybe even introduce a conjunction again. Since we have impossible worlds around, conjunction introduction or elimination might fail for this particular case. Moreover, even if we accept that their conjunction is true, we might be in a non-explosive world.

5. Answering Objections

We close by discussing, in this order: unreliable narration; the assumption of the plausibility ordering; objectivity of truth in fiction; and the objection by Bonomi and Zucchi [2003].

While unreliable narration and non-literal speech may pose complications in our framework, too, these issues have troubling counterparts in the general modelling of communication of real agents and are thus not really a problem *specific* to narrators, or depiction in movies. Consider a coin tossing game with three players: Alicia, Bono, and Circe (a, b, c). Suppose that b is tossing the coin and takes a peek. a and c see this, but they do not peek. b says, 'The coin shows Heads.' How do they know that b is reliable? They usually simply assume it. If they have evidence that he's not, they might not upgrade their belief system or they might, but index this particular move somehow. The communication between narrator and reader, it seems to us, is similar to real communication: solutions to the coin-toss case might be transferable to the case of fiction.

Next, the plausibility ordering: just like the Lewisian similarity ordering between possible worlds, this is rather vague. Moreover, assuming that agents come equipped with such an order is a fairly strong move. Here are some responses concerning this objection. First, it is fairly standard in contemporary epistemic-doxastic logic to assume such orderings. If one objects to plausibility orderings in general, one has to object to the whole mainstream approach to belief revision of epistemic-doxastic logic, too. Second, in the debate about truth in fiction, everyone *is* already implicitly assuming some plausibility ordering on worlds. Any analysis of truth in fiction is evaluated against our intuitions about what is true in a fiction (see the objections above against Analyses 1 and 2). But such intuitions are never justified when used to evaluate the analysis. We entertain certain scenarios, where we pretend that the story obtains, and then wonder whether it is intuitively true in this scenario that A or not. But how can we say that it is intuitively not true in the Holmes stories that Trump wins the election in 2016? Or that it is intuitively true in the Holmes stories that Holmes, when reasoning deductively, is reasoning classically? It is because we, as interpreters, consider one interpretation more *plausible* than another. (Closer to Lewis: we, as interpreters, believe that the community of origin of *f* considered one interpretation more plausible than another.) Hence, we claim, this appeal to intuition in the debate is an appeal to plausibility among interpretations. We model this as a plausibility ordering on worlds.

This leads to another possible objection: our analysis makes truth in fiction too community-dependent. Lewis, but also Currie and others, seems to assume that there is a fact of the matter to what's true in a story and that an analysis of truth in fiction should reveal this. Folde [2015] argues that correct interpretations give us epistemic access to what's true in the fiction, and correct interpretations are grounded in what's true in the fiction.

First, though, it's not entirely clear whether the assumption that there's a fact of the matter to what's true in a story is right. In particular, it is unclear whether there is a fact of the matter to what is the implicit content of a story. Some conclusions about what's true in a fiction, as noted by Hanley [2004], are *probably* true or supported by inductive arguments. But we can most certainly go to the literary studies department and find many works of fiction for which there are competing interpretations giving us competing claims about what's true in the fiction. It might be unfeasible to determine which of the interpretations is actually tracking the facts of the matter, assuming that there is such a thing. So, even if there is a fact of the matter concerning the implicit content of a story, it's only accessible through our interpretation and our background information.

The open issue here can be formulated as a dilemma:[11] either, strictly speaking, nothing is true in a fiction independent of interpretation (ours, or the relevant community's), and rather, while reading fiction, we provide an interpretation: by doing so, we *fix* what's true in the story. Or, if fictions instead come with a fact of the matter fixed in the corresponding fictional worlds, then while interpreting we try to *discover* these facts.

Neither option seems satisfying. The former treats truth in fiction as too dependent on the community of interpretation, and fails to fit the phenomenology that we have when reading, which is that we are discovering something about the fictional world(s). The latter leads to epistemic difficulties. How could we ever know that we have discovered the facts with our interpretation, given that there is at least one competing, but equally well justified, interpretation contradicting ours?

We think that our analysis provides a good model of literary theory practice and of how truth in fiction is approached there. If there's disagreement in the community of interpretation, then we can account for this disagreement, because we refer to the community of interpretation in the world of evaluation, and not to that of origin. It's simply undecided whether it's true in the story that A or $\neg A$ because the most plausible worlds are going to contain A-worlds *and* $\neg A$-worlds.

The worry from Bonomi and Zucchi [2003] was with the community of origin's common beliefs being imported into the story. As the group ordering is based on the hierarchy of agents of the community, depending on who gets the highest ranking, we might block their objection. The easiest way is to give highest authority to the author. Another way is to consider, as we did, the community of interpretation, instead of the community of origin. It will then revise with the fiction's explicit content and can, before engaging with the fiction, consider genre conventions or other pragmatic criteria to reorder the worlds. In particular, *some* of the common beliefs of the community of origin might be considered, and used to reorder the worlds that the community of interpretation considers. However, the latter might be resisting the import of certain moral judgments of the community of origin.

Thus, our view may also allow us to model imaginative resistance [Gendler 2000]. We don't have the room to discuss imaginative resistance in detail, but here are some thoughts, by way of conclusion. Gendler believes that we are unwilling to make-believe that eating babies is morally right, due to an exportation principle: if we make-believe that eating babies is morally right, we are seemingly justified in exporting this as a moral truth *simpliciter*. One may model this, in our model, by adding a moral modality [MOK] ('It is morally OK that'), designating the actual world $@ \in N$, and stipulating that if $\rho_@(\neg[\mathrm{MOK}]\,A)1$ and $\rho_w([\mathrm{MOK}]A)1$ then w is always ranked below any f-world where $\neg[\mathrm{MOK}]\,A$.[12]

Funding

Franz Berto's research was funded by the European Research Council (ERC CoG), Consolidator grant no. 681404, 'The Logic of Conceivability'. Christopher Badura's research was funded by the Ruhr University Research School PLUS, funded by Germany's Excellence Initiative [DFG GSC 98/3].

[11] Thanks to Nathan Wildman for this neat formulation.

[12] We are grateful to Aybüke Özgün, Graham Priest, Tom Schoonen, and two anonymous referees for helpful comments and remarks.

References

Arrow, K.J. 1950. A Difficulty in the Concept of Social Welfare, *Journal of Political Economy* 58/4: 328–46.

Baltag, A. and S. Smets 2011. Keep Changing Your Beliefs, Aiming for the Truth, *Erkenntnis* 75/2: 255–70.

Barthes, R. 1977. The Death of the Author, in *Image—Music—Text*, ed. S. Heath, London: Fontana Press: 142–9.

Berto, F. 2013. Impossible Worlds, in *Stanford Encyclopedia of Philosophy (Winter 2013 Edition)*, ed. Edward N. Zalta, URL = https://plato.stanford.edu/archives/win2013/entries/impossible-worlds

Bonomi, A. and S. Zucchi 2003. A Pragmatic Framework for Truth in Fiction, *Dialectica* 57/2: 103–20.

Bourne, C. and E.C. Caddick 2016. *Time in Fiction*, Oxford: Oxford University Press.

Currie, G. 1990. *The Nature of Fiction*, Cambridge: Cambridge University Press.

Folde, C. 2011. *Was ist Wahrheit in Fiktion? Eine Auseinandersetzung mit David Lewis*, Master's Dissertation, Universität Hamburg.

Folde, C. 2015. Grounding Interpretation, *British Journal of Aesthetics* 55/3: 361–74.

Friend, S. 2012. Fiction as a Genre, *Proceedings of the Aristotelian Society* 112/2: 179–209.

Gendler, T.S. 2000. The Puzzle of Imaginative Resistance, *The Journal of Philosophy* 97/2: 55–81.

Gendler, T.S. 2011. Is Dumbledore Gay? *The Philosophers' Magazine* 52/1: 94–7.

Girard, P. and K. Tanaka 2016. Paraconsistent Dynamics, *Synthese* 193/1: 1–14.

Grove, A. 1988. Two Modellings for Theory Change, *Journal of Philosophical Logic* 17/2: 157–170.

Hanley, R. 2004. As Good as It Gets: Lewis on Truth in Fiction, *Australasian Journal of Philosophy* 82/1: 112–28.

Heyd, T. 2006. Understanding and Handling Unreliable Narratives: A Pragmatic Model and Method, *Semiotica* 2006/162: 217–43.

Jago, M. 2014. *The Impossible: An Essay on Hyperintensionality*, Oxford: Oxford University Press.

Kiourti, I.G. 2010. *Real Impossible Worlds: The Bounds of Possibility*, Ph.D. Dissertation, University of St. Andrews.

Klassen, T., H.J. Levesque and S.A. McIlraith 2017. Towards Representing What Readers of Fiction Believe, in *Proceedings of the Thirteenth International Symposium on Commonsense Reasoning, COMMONSENSE 2017, London, UK, November 6–8, 2017*, CEUR Conference Proceedings, 2052, ed. A.S. Gordon, R. Miller, and G. Turán, CEUR-WS.org, URL = <https://ceur-ws.org/Vol-2052/paper12.pdf>.

Köppe, T. and T. Kindt 2011. Unreliable Narration with a Narrator and Without, *Journal of Literary Theory* 5/1: 81–93.

Leitgeb, H. and K. Segerberg 2007. Dynamic Doxastic Logic: Why, How, and Where To? *Synthese* 155/2: 167–90.

Lewis, D. 1978. Truth in Fiction, *American Philosophical Quarterly* 15/1: 37–46.

Lewis, D. 1983. Postscript to Truth in Fiction, in his *Philosophical Papers: Volume 1*, New York: Oxford University Press: 276–80.

Löwe, B. 2010. Comparing Formal Frameworks of Narrative Structures, in *Computational Models of Narrative. Papers from the 2010 AAAI Fall Symposium* [AAAI Technical Report FS-10-04], ed. Mark Finlayson, Menlo Park, CA: AAAI Press 2010: 45–6.

Matravers, D. 2014. *Fiction and Narrative*, Oxford: Oxford University Press.

Nichols, S. and S. Stich 2003. *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*, Oxford: Clarendon Press.

Nolan, D. 1997. Impossible Worlds: A Modest Approach, *Notre Dame Journal of Formal Logic* 38/4: 535–72.

Nolan, D. 2007. A Consistent Reading of Sylvan's Box, *The Philosophical Quarterly* 57/229: 667–73.

Nolan, D. 2013. Impossible Worlds, *Philosophy Compass*, 8/4: 360–72.

Priest, G. 1997. Sylvan's Box: A Short Story and Ten Morals, *Notre Dame Journal of Formal Logic* 38/4: 573–82.

Priest, G. 2008. *An Introduction to Non-Classical Logic: From If to Is*, 2nd edn, Cambridge: Cambridge University Press.

Proudfoot, D. 2006. Possible Worlds Semantics and Fiction, *Journal of Philosophical Logic* 35/1: 9–40.

Rapaport, W.J. and S.C. Shapiro 1995. Cognition and Fiction, in *Deixis in Narrative: A Cognitive Science Perspective*, ed. J.F. Duchan, G.A. Bruder, and L.E. Hewitt, Hillsdale, NJ and Hove, UK: Lawrence Erlbaum: 107–28.

Riggan, W.E. 1981. *Picaros, Madmen, Naïfs, and Clowns: The Unreliable First-Person Narrator*, Norman, OK: University of Oklahoma Press.

Sainsbury, R.M. 2010. *Fiction and Fictionalism*, London: Routledge.

Searle, J.R. 1975. The Logical Status of Fictional Discourse, *New Literary History* 6/2: 319–32.

Segerberg, K. 1995. Belief Revision from the Point of View of Doxastic Logic, *Logic Journal of IGPL* 3/4: 535–53.

Segerberg, K. 2001. The Basic Dynamic Doxastic Logic of AGM, *Frontiers in Belief Revision* APLS 22: 57–84.

van Benthem, J. 2007. Dynamic Logic for Belief Revision, *Journal of Applied Non-Classical Logics* 17/2: 129–55.

van Ditmarsch, H.P. 2005. Prolegomena to Dynamic Logic for Belief Revision, *Synthese* 142/2: 229–75.

Walton, K.L. 1990. *Mimesis as Make-Believe: On the Foundations of the Representational Arts*, Cambridge, MA: Harvard University Press.