Towards Representing What Readers of Fiction Believe

Toryn Q. Klassen and **Hector J. Levesque** and **Sheila A. McIlraith**
Department of Computer Science
University of Toronto
{toryn,hector,sheila}@cs.toronto.edu

Abstract

Despite the extensive literature on the problem of story understanding, there has been little focus on formally representing some forms of knowledge that are specific to stories, such as how the reader expects information to be presented over the course of reading. To illustrate, the reader of a mystery story may expect to eventually find out who is guilty, and also that the author may first try to mislead them about who is guilty. We propose *literary logic*, a formalism based on work by Friedman and Halpern for reasoning about dynamic systems, and apply it in representing this sort of knowledge. We also consider issues relating to carrying over world knowledge into fiction, and knowledge of genre conventions.

1 Introduction

Story understanding is a long-standing problem in artificial intelligence, with notable early work from the 1970s (Charniak 1972; Schank and Abelson 1977). McCarthy (1990), in a memo originally from 1976, pointed out that stories raise problems for commonsense reasoning. Research has continued, and recent years have seen a proliferation of corpora of stories in various mediums with accompanying questions for machine learning purposes, including MCTest (Richardson, Burges, and Renshaw 2013), ROCStories (Mostafazadeh et al. 2016), MovieQA (Tapaswi et al. 2016), and COMICS (Iyyer et al. 2017). In this paper, we are concerned with the task of a reader answering questions about a story after reading (some prefix of) it.[1] We propose a logic for the purpose of determining how the reader would answer such questions based on their various types of background knowledge.

Much of the work on story understanding has focused on the "world knowledge" needed to understand stories. For example, Charniak (1972) devoted a chapter to how knowledge about piggy banks can be used in understanding passages about them. In representing stereotypical events, scripts (Schank and Abelson 1977) also encode world knowledge, like that tips are given at restaurants after eating.

However, there are other forms of knowledge that are also relevant. Diakidoy et al. (2014) suggested that readers have "story knowledge" such as expectations that characters' efforts would meet with complications, but they did not try to represent that in their argumentation-based approach to story understanding. Some information of that sort could be represented in a story grammar (Rumelhart 1975). Charniak and Goldman (1989) pointed out the significance of readers assuming that mentioned objects are going to be relevant. In work on using abduction to interpret text (Hobbs, Stickel, and Martin 1993), it's been suggested that the abductive explanations might refer to such things as authors' plans. Despite interest in interpreting literature (Hobbs 1990), this has not been much focused on in the context of stories. We may note that if scripts are learned from corpora, as by Chambers and Jurafsky (2009), they probably end up also capturing information about what events authors find noteworthy. Chaturvedi, Peng, and Roth (2017) consider several forms of knowledge in trying to predict the correct ending of a story, including knowledge of patterns of sentiment in stories.

Forms of knowledge which have not been explored in much depth include what the reader believes that they will come to learn from reading (parts of) the story, and what the reader thinks the author will try to make them believe over time. For example, the reader may believe that they will learn from reading a mystery who was guilty, but that the author will try to make them believe at some time that an innocent character is guilty. Or the reader may believe that if they haven't been told the main character's eye color by halfway through a book, they'll never find it out. In this paper, we apply an approach to modelling belief and time to this sort of representational problem. We also focus specifically on fiction, unlike most AI story understanding research.

Applying world knowledge to fiction is more complicated than to non-fiction. In the philosophical literature, there has been substantial work on defining "truth in fiction". Lewis (1978) noted the phenomenon of *carry-over*, that "factual premisses [...] may carry over into the fiction, not because there is anything explicit in the fiction to make them true, but rather because there is nothing to make them false" (p. 42). To use an example of his, we may assume that Sherlock Holmes does not have a third nostril. Lewis offered multiple definitions of truth in fiction; his "Analysis I" said that

> A sentence of the form "In the fiction f, ϕ" is non-vacuously true iff some world where f is told as known fact and ϕ is true differs less from our actual world, on balance, than does any world where f is told as known fact and ϕ is not true.

[1] We will not be considering summarization or further tasks that have been suggested as part of story understanding (Michael 2013).

His "Analysis II" was similar but instead of considering differences from the actual world, considered differences from the worlds where the common beliefs of the fiction's community of origin were true. Others have used similar ideas, e.g. Walton (1990) had his "Reality Principle" and "Mutual Belief Principle" which roughly correspond to Lewis's analyses, though Walton regarded them only as rules of thumb. Genre information – e.g., about time travel (Morgenstern 2014), or that dragons breathe fire – is another sort of knowledge, which is unclear how to incorporate into these sorts of definitions; perhaps the most detailed approach attempting to do so was given by Bonomi and Zucchi (2003).

These philosophical approaches were not fully formalized and expressed in logic. The logics designed by philosophers for dealing with fiction (Woods 1974; Heintz 1979) have usually focused on other issues, like handling inconsistent stories (which we will not be addressing in this paper).

The formal logic we present in this paper, which we call *literary logic* ($\mathcal{LL}$), is a variant of the logic used by Friedman and Halpern (1999) to model belief revision in dynamical systems. We argue that $\mathcal{LL}$ can be used to represent various forms of knowledge relevant to story understanding. We focus on two main issues: representing reader's expectations about stories (which may take into account genre-specific information), and the carry-over of world knowledge and its interaction with genre knowledge (e.g. about dragons). Literary logic provides temporal features that we apply to the first issue (though they may also have a role to play with respect to the second), and non-monotonic aspects that are useful for both. The outline of this paper is as follows. Section 2 describes the syntax and semantics of literary logic, and notes some of its properties. Section 3 shows how the question-answering task can be formalized, describing how we can make use of abnormality predicates (McCarthy 1986) in specifying the reader's initial epistemic state. Section 4 formalizes some examples of reader knowledge: we consider carry-over (and incorporating genre knowledge) in section 4.1, and then expectations about mystery stories in section 4.2. Section 5 discusses related work, and section 6 concludes with a discussion of future work.

2 Literary logic

This section describes the language $\mathcal{LL}$, which is closely based on the logic of Friedman and Halpern, which provided for modelling the accessibility and plausibility of possible worlds over time. The major differences include that $\mathcal{LL}$ is first-order, is evaluated with respect to finite rather than infinite timelines (because stories are finite and are read in finite time), and includes the complete set of past and future temporal operators from Lichtenstein, Pnueli, and Zuck (1985). $\mathcal{LL}$ describes the beliefs of a reader over time as they read a *discourse*, a sequence of logical sentences representing a story, one sentence per time step.

A very visible feature of $\mathcal{LL}$ (that is mostly just for clarity of presentation) is that we have two sorts of predicates, "real" and "imaginary" ones, and have a special "in imagination" operator $\mathbf{I}$. The idea is that real predicates describe properties in the real world (including "literary" properties, like the genre of the story being read) while imaginary predicates (that hold only "in imagination") describe properties that apply within the world of the story being read. The reader's beliefs about the extensions of both sorts of predicates can change over the course of reading.

2.1 Syntax

The syntax of $\mathcal{LL}$ involves both terms and predicates.

A *term* is either a *standard name* or a *variable*. There is a countably infinite set $\mathcal{N} = \{{}^{\#}1, {}^{\#}2, {}^{\#}3, \dots\}$ of standard names. Intuitively, these stand for all the objects that we may want to refer to, including not just real-life things like piggy banks, but also theoretical literary concepts like what Van Inwagen (1977) called "creatures of fiction", like the character Sherlock Holmes or his pipe. There also is a countably infinite set of variables. Note that the logic does not have constants or function symbols, though the standard names can be thought of as constant symbols that satisfy the unique name assumption and an infinitary version of domain closure. For a discussion of why standard names are useful, see Levesque and Lakemeyer (2000, section 2.2).

As previously indicated, there are two (non-empty) sets of *predicate symbols*, the real Φ_r and the imaginary Φ_f (which do not have to be disjoint). Each predicate P from either set has an arity, $\mathsf{ar}(P)$, which is the number of terms that it takes as arguments. So, to give a typical example, there could be a real unary predicate Rabbit that indicates its argument is a rabbit in reality, and an imaginary unary predicate also called Rabbit that indicates its argument is a rabbit in the world of the story under consideration. The set of real predicates would also typically include predicates to express literary propositions; for example, there could be a 0-ary real predicate $\mathit{FantasyGenre}$ which would indicate that the story was in the fantasy genre.

A *real atom* is a string of the form $P(t_1, \dots, t_k)$, where $P \in \Phi_r$, $k = \mathsf{ar}(P)$, and $t_1, \dots, t_k$ are terms. Similarly, an *imaginary atom* is a string of the form $Q(t_1, \dots, t_k)$, where $Q \in \Phi_f$. We will say that an atom is *ground* if no variables appear in it. We will assume that there is a unary predicate $\mathit{Mentioned} \in \Phi_r$, which we will later give the special meaning of picking out those standard names that appear within the discourse.

The formulas of $\mathcal{LL}$ are the expressions of the form ϕ generated by the grammar below, where P is a real atom, Q is an imaginary atom, x is a variable, and t_1 and t_2 are terms.

$$\begin{aligned} \alpha &\coloneqq Q \mid \neg\alpha \mid (\alpha \wedge \alpha) \mid (t_1 = t_2) \mid \exists x(\alpha) \\ \phi &\coloneqq P \mid \neg\phi \mid (\phi \wedge \phi) \mid (t_1 = t_2) \mid \exists x(\phi) \mid \mathbf{I}\alpha \mid \\ &\qquad \mathbf{D}\alpha \mid \bigcirc\phi \mid \bullet\phi \mid \phi \,\mathbf{U}\, \phi \mid \phi \,\mathbf{S}\, \phi \mid \phi \,\square\!\!\rightarrow \phi \end{aligned}$$

We will also be talking about α-type formulas, which are expressions of the form α generated by the grammar (though the ϕ-type formulas are what we will mean when we refer to $\mathcal{LL}$ formulas). A variable x appearing in a (α- or ϕ-type) formula is said to be *free* if it does not appear within a subformula of the form $\exists x(\phi)$, and a formula with no free variables is called a *sentence*. The use to which we will put α-type sentences is that the discourse being read is a sequence of α-type sentences, which describe the world of the story.

An $\mathcal{LL}$ sentence (that is, a ϕ-type sentence) describes the real world, and can include modal operators to describe the reader's beliefs over time.

The operators $\neg, \wedge, \exists$, and $=$ are familiar from first-order logic, and we can use them to define abbreviations like $\vee, \supset$, $\equiv$, and $\forall$ in the usual ways. It's convenient to have a symbol $\top$ that always takes a true truth value; let $\top := \forall x(x = x)$.

We will read $\mathbf{I}\alpha$ as "α is imagined" or "α is true in imagination" (though the $\mathbf{I}$ operator serves a technical function and is not intended to formalize a commonsense notion of imagination). We will read $\mathbf{D}\alpha$ as "The last sentence read of the discourse was α".

The operators $\bigcirc$ ("next"), $\bullet$ ("previous"), $\mathbf{U}$ ("until"), and $\mathbf{S}$ ("since") are standard temporal logic operators.[2] They describe time for the reader, who reads one sentence of the discourse each time step. We can define further temporal operators in the usual ways: $\Diamond$ ("eventually") by $\Diamond\phi := \top\,\mathbf{U}\,\phi$, $\Box$ ("always in the future") by $\Box\phi := \neg\Diamond\neg\phi$, $\blacklozenge$ ("sometime in the past") by $\blacklozenge\phi := \top\;\mathbf{S}\;\phi$, and $\blacksquare$ ("always in the past") by $\blacksquare\phi := \neg\blacklozenge\neg\phi$. We can also define an operator $\blacktriangleright\!\!|$ ("after reading") by $\blacktriangleright\!\!|\,\phi := \Diamond(\phi \wedge \neg\bigcirc\top)$, so $\blacktriangleright\!\!|\,\phi$ means that ϕ is true at the final time (i.e., when the entire story has been read), and $|\!\!\blacktriangleleft$ ("initially") by $|\!\!\blacktriangleleft\phi := \blacklozenge(\phi \wedge \neg\bullet\top)$, so that $|\!\!\blacktriangleleft\varphi$ means ϕ is true at time 0 (the initial time).

The formula $\phi_1 \mathrel{\Box\!\!\rightarrow} \phi_2$ means that ϕ_2 is true in all the *most plausible* accessible worlds in which ϕ_1 is true. We could follow Friedman and Halpern in defining a belief operator $\mathbf{B}$ with the abbreviation $\mathbf{B}\phi := \top \mathrel{\Box\!\!\rightarrow} \phi$ (that is, ϕ is believed if it is true in all the most plausible accessible worlds), but instead let us give a more general definition: if ψ is any sentence, let

$$\mathbf{B}_\psi\phi := (|\!\!\blacktriangleleft\psi) \mathrel{\Box\!\!\rightarrow} \phi. \tag{1}$$

We can think of $\mathbf{B}_\psi\phi$ as indicating that ϕ is believed by an agent who initially considers it impossible that ψ is false. We will call ψ the *knowledge base* (or KB) of the agent (though ψ is not necessarily true). Note that ψ cannot include the $\mathbf{B}_\psi$ operator, for then $\mathbf{B}_\psi\phi$ would not expand to a finite sentence, but ψ can contain $\mathbf{B}_{\psi'}$ for a suitable different sentence ψ'. We also define a "knowledge" operator $\mathbf{K}_\psi$ by

$$\mathbf{K}_\psi\phi := ((|\!\!\blacktriangleleft\psi) \wedge \neg\phi) \mathrel{\Box\!\!\rightarrow} \neg\top. \tag{2}$$

The result is that $\mathbf{K}_\psi\phi$ is true if the agent with knowledge base ψ considers it impossible that ϕ is false.

We define (a subjective version of) *fictional truth* to be what the reader, after reading the entire story, believes is imagined:

$$\mathbf{F}_\psi\alpha := \blacktriangleright\!\!|\,\mathbf{B}_\psi\mathbf{I}\alpha. \tag{3}$$

We can read $\mathbf{F}_\psi\alpha$ as saying that α is (subjectively) fictionally true. The reason why we want to consider the final time (and so use the $\blacktriangleright\!\!|$ operator) is that fictional truth is determined by the story as a whole, which has only been fully consumed at the final time.

Furthermore, we define $[\alpha]\phi$ by $[\alpha]\phi := (\bigcirc\mathbf{D}\alpha \supset \bigcirc\phi)$. So $[\alpha]\phi$ says that ϕ is true after reading α (provided that the next sentence is actually α). We will abbreviate sequences of such operators with $[\alpha_1; \ldots; \alpha_k]\phi := [\alpha_1]\cdots[\alpha_k]\phi$. So, e.g., $[\alpha_1; \alpha_2]\phi$ abbreviates $(\bigcirc\mathbf{D}\alpha_1 \supset \bigcirc(\bigcirc\mathbf{D}\alpha_2 \supset \bigcirc\phi))$.

[2] Our "next" and "previous" operators are the "strong" versions.

2.2 Semantics

The grammar provides two types of formulas, denoted by α and ϕ. While our goal in this section is to define satisfaction and validity with respect to ϕ-type sentences, let us first define a satisfaction relation $\models$ that specifies when an α-type sentence is satisfied by an *interpretation* π (which we take to be a set of imaginary ground atoms). We will use the notation $\alpha[x/c]$ to indicate the formula obtained by replacing all free occurrences of the variable x in α by $c \in \mathcal{N}$.

1. $\pi \models Q(c_1, \ldots, c_k)$ iff $Q(c_1, \ldots, c_k) \in \pi$
2. $\pi \models \neg\alpha$ iff $\pi \not\models \alpha$
3. $\pi \models (\alpha_1 \wedge \alpha_2)$ iff $\pi \models \alpha_1$ and $\pi \models \alpha_2$
4. $\pi \models (c_1 = c_2)$ iff c_1 and c_2 are identical names
5. $\pi \models \exists x(\alpha)$ iff $\pi \models \alpha[x/c]$ for some $c \in \mathcal{N}$

This is just an established way of giving the semantics of a first-order logic with substitutional quantification (Levesque and Lakemeyer 2000). Now, let us make some definitions.

Definition 1 (discourse)**.** A discourse is a finite sequence of α-type sentences, ending with End, a special sentence not appearing earlier (we do not have to introduce a new symbol for this; we can just take $\mathsf{End} = \top$).

Definition 2 (complex world)**.** A (complex) world is a tuple $w = \langle w_r, w_f, w_d\rangle$ where w_r is a set of real ground atoms, w_f is a set of imaginary ground atoms, and w_d is a discourse s.t. (1) if $w_d(i) = \alpha$ for any i then $w_f \models \alpha$, and (2) iff $c \in \mathcal{N}$ appears in a sentence of w_d, then $\mathit{Mentioned}(c) \in w_r$.

Intuitively, w_r is the set of all real ground atoms that are true in the world w, w_f is the set of all imaginary ground atoms that are true in w, and w_d is a formal representation of the story that is told in w. Note that w_f represents one way of "completing" the fictional world in a way compatible with the story being told. What w_f makes true is not the same as what is fictionally true (as determined by a $\mathbf{F}_\psi$ operator).

The sentences of a discourse, unlike those of a natural language story, are not indexical relative to the "current" time within the story. So, for example, the rather trivial story "John picked up a block. Then he put it back down." could get encoded (in a style based after Maslan, Roemmele, and Gordon (2015)) as the following discourse: $\langle \mathit{John}(\#1) \wedge \mathit{Block}(\#2) \wedge \mathit{Pickup}(\#1, \#2, \#3),\ \ \mathit{Precedes}(\#3, \#4) \wedge \mathit{Putdown}(\#1, \#2, \#4),\ \mathsf{End}\rangle$. The last arguments to Pickup and $\mathit{Putdown}$ are meant to be the names of event instances, so e.g. $\mathit{Pickup}(\#1, \#2, \#3)$ says that $\#3$ is an event in which $\#1$ picked up $\#2$, and $\mathit{Precedes}$ expresses the events' temporal ordering (with respect to time within the story, not time for the reader). The point here however is not the specific way these sentences represent time, but that they are like what Quine (1968) called "eternal sentences" in that their truth does not depend on their time of evaluation. We will also expect a discourse to usually provide standard names for relevant objects and events, as our example did.

In order to provide semantics for the $\mathrel{\Box\!\!\rightarrow}$ operator, we need a way to represent plausibility. Friedman and Halpern did so using the very general notion of a *plausibility space*; we will use what can be considered a special case of that, a version of the popular "system of spheres" representation

(Lewis 1973; Grove 1988; Bonomi and Zucchi 2003). Below we will use $\mathcal{W}$ to denote the set of all complex worlds.

Definition 3 (system of spheres)**.** A system of spheres is a set $\mathcal{S}$ of subsets ("spheres") of $\mathcal{W}$ such that (1) for any two spheres $U \in \mathcal{S}$ and $V \in \mathcal{S}$, either $U \subseteq V$ or $V \subseteq U$, (2) for any non-empty set $V \subseteq \mathcal{W}$, there is a $\subseteq$-minimal sphere C such that $C \cap V \neq \emptyset$, and (3) $\mathcal{W} \in \mathcal{S}$.

A system of spheres can also be thought of as a total preorder $\preceq$ on worlds, where $w \preceq v$ ("w is at least as plausible as v") if every sphere containing v also contains w. Every system of spheres has a "central" sphere (the $\subseteq$-minimal sphere C such that $C \cap \mathcal{W} \neq \emptyset$) containing the $\preceq$-minimal (most plausible) worlds.

The $\Box\!\!\rightarrow$ operator depends not just on the plausibility of worlds, but on which worlds are (currently) *accessible*.

Definition 4. For b a non-negative integer, the *accessibility relation at time* b, $\sim_b \subseteq \mathcal{W} \times \mathcal{W}$, is given by $w \sim_b v$ iff $|w_d| \geq b$, $|v_d| \geq b$, and $w_d(i) = v_d(i)$ for $1 \leq i \leq b$.

Intuitively, at time b the reader will not consider possible any world with a discourse not starting with the same b sentences they have read so far. For a world w with $|w_d| \geq b$ we may use the notation $[w]_{\sim_b} := \{v \in \mathcal{W} : w \sim_b v\}$. That is, $[w]_{\sim_b}$ is the set of worlds accessible from w at time b.

The satisfaction of a literary logic sentence ϕ is given relative to a system of spheres $\preceq$, a world $w = \langle w_r, w_f, w_d \rangle \in \mathcal{W}$, and a time $b \in \{0, 1, \ldots, n\}$, where $n = |w_d|$, the length of the discourse w_d. The recursive rules for when $\preceq, w, b$ satisfy ϕ, written $\preceq, w, b \Vdash \phi$, are given below:

1. $\preceq, w, b \Vdash P(c_1, \ldots, c_k)$ iff $P(c_1, \ldots, c_k) \in w_r$
2. $\preceq, w, b \Vdash \neg\phi$ iff $\preceq, w, b \nVdash \phi$
3. $\preceq, w, b \Vdash (\phi_1 \wedge \phi_2)$ iff $\preceq, w, b \Vdash \phi_1$ and $\preceq, w, b \Vdash \phi_2$
4. $\preceq, w, b \Vdash (c_1 = c_2)$ iff c_1 and c_2 are identical names
5. $\preceq, w, b \Vdash \exists x(\phi)$ iff $\preceq, w, b \Vdash \phi[x/c]$ for some $c \in \mathcal{N}$
6. $\preceq, w, b \Vdash \mathbf{I}\alpha$ iff $w_f \models \alpha$
7. $\preceq, w, b \Vdash \mathbf{D}\alpha$ iff $b > 0$ and $w_d(b) = \alpha$
8. $\preceq, w, b \Vdash \bigcirc\phi$ iff $b < n$ and $\preceq, w, b+1 \Vdash \phi$
9. $\preceq, w, b \Vdash \bullet\phi$ iff $b > 0$ and $\preceq, w, b-1 \Vdash \phi$
10. $\preceq, w, b \Vdash \phi_1 \,\mathbf{U}\, \phi_2$ iff $\preceq, w, j \Vdash \phi_2$ for some j such that $b \leq j \leq n$ and $\preceq, w, k \Vdash \phi_1$ for all k s.t. $b \leq k < j$
11. $\preceq, w, b \Vdash \phi_1 \,\mathbf{S}\, \phi_2$ iff $\preceq, w, j \Vdash \phi_2$ for some j such that $0 \leq j \leq b$ and $\preceq, w, k \Vdash \phi_1$ for all k s.t. $j < k \leq b$
12. $\preceq, w, b \Vdash \phi_1 \Box\!\!\rightarrow \phi_2$ iff $\preceq, v, b \Vdash \phi_2$ for every $v \in \min_{\preceq}\{v \in [w]_{\sim_b} : \preceq, v, b \Vdash \phi_1\}$

We will write $\preceq, w \Vdash \phi$ if $\preceq, w, 0 \Vdash \phi$. We will write $\Vdash \phi$ ("ϕ is valid") if $\preceq, w \Vdash \phi$ for every system of spheres $\preceq$ and world w.

2.3 Properties

To understand the $\mathbf{B}_\psi$ operator, it is helpful to introduce another accessibility relation, $\sim_b^\psi \subseteq \mathcal{W} \times \mathcal{W}$, where b is a time and ψ a sentence.

Definition 5. Given a system of spheres $\preceq$, a time b, and sentence ψ, define $w \sim_b^\psi v$ iff $w \sim_b v$ and $\preceq, v, 0 \Vdash \psi$.

Intuitively, $w \sim_b^\psi v$ if at world w and time b, the reader with knowledge base ψ considers world v possible.

Observation 1. $\preceq, w, b \Vdash \mathbf{B}_\psi\phi$ iff $\preceq, v, b \Vdash \phi$ for every $v \in \min_{\preceq}([w]_{\sim_b^\psi})$.

$\mathbf{B}_\psi$ can be shown to be a K45 operator (supporting positive and negative introspection). There is also remembrance of past beliefs, e.g. we have $\Vdash \bullet\mathbf{B}_\psi\phi \supset \mathbf{B}_\psi\bullet\mathbf{B}_\psi\phi$.

Observation 2. While $\Box(\mathbf{I}(\alpha_1 \vee \alpha_2) \supset (\mathbf{I}\alpha_1 \vee \mathbf{I}\alpha_2))$ and $\Box(\mathbf{I}(\exists x\alpha) \supset \exists x\mathbf{I}\alpha)$ are valid for any α_1 and α_2, $\mathbf{F}_\psi(\alpha_1 \vee \alpha_2) \supset (\mathbf{F}_\psi\alpha_1 \vee \mathbf{F}_\psi\alpha_2)$ and $\mathbf{F}_\psi(\exists x\alpha_1) \supset \exists x\mathbf{F}_\psi\alpha_1$ are not (assuming for the last that some $Q \in \Phi_f$ has nonzero arity).

Note that if Observation 2 did not hold the behavior of the $\mathbf{F}_\psi$ operator would contradict the generally accepted idea that fiction is incomplete (Doležel 1995) and so there is no answer to the question of, for example, exactly how many children Lady Macbeth had in Shakespeare's *Macbeth* (Wolterstorff 1976). As McCarthy (1990) wrote, "In a made-up story, questions about middle names or what year the story occurred in do not necessarily have an answer".

Observation 3. $\Vdash \Box(\mathbf{D}\alpha \supset \mathbf{F}_\psi\alpha)$. That is, whatever the discourse includes is fictionally true for any reader.

Note this means we cannot encode metaphorical language. Also, Walton (1990, §4.5) raised philosophical questions on how literally some other aspects of stories should be taken, such as whether Shakespearean characters are really fictionally uttering the poetic speeches attributed to them (or, more simply, whether characters are really speaking English in English-language stories). Our formalism does not offer a choice in how to answer that.

3 Applying $\mathcal{LL}$ to question-answering

We want to formalize how a reader would answer questions after reading (a prefix of) a story. As part of this formalization, we want to specify (within the language) not just what the reader initially believes, but what things the reader initially considers more plausible than others (so as to determine exactly how the reader's beliefs evolve in response to reading). In $\mathcal{LL}$, following some previous work on belief revision (Friedman and Halpern 1999; Shapiro et al. 2011), the plausibility of worlds does not actually change over time, but only the accessibility relation. That nonetheless suffices to allow whether a proposition is believed to change back and forth over time (see (Shapiro et al. 2011, section 6)). This suggests we can fix one system of spheres to always use, and just set the initial accessibility relation appropriately (which the ψ in the $\mathbf{B}_\psi$ operator has the effect of doing). In this section, we will define the '$\Vdash\!\sim$' relation, the analogue of $\Vdash$ for a particular fixed system of spheres. Then question-answering can be done by determining which expressions of the form $\Vdash\!\sim [\alpha_1; \ldots; \alpha_k]\mathbf{B}_\psi\phi$ hold, i.e. what a reader with a KB ψ of background knowledge (of possibly various types) believes after reading the first k sentences of the discourse.

To define the specific system of spheres, we will apply the idea of circumscription (McCarthy 1986) and have the plausibility of worlds be inversely related to the sizes of the extensions of distinguished "abnormality" predicates. Suppose that we have a finite set of abnormality predicates, each with an associated priority (a positive integer). If Ab is a k-ary abnormality predicate of priority i, we will say that $Ab(c_1, \ldots, c_k)$ is a priority i ground atom. For a world w, let $C_i(w) \in \{0, 1, 2, \ldots\} \cup \{\infty\}$ be the sum of numbers of priority i ground atoms from w_r and w_f. Let the partial order $\prec_{\mathsf{CIRC}} \subseteq \mathcal{W} \times \mathcal{W}$ be defined by $w \prec_{\mathsf{CIRC}} v$ if there is some i for which $C_i(w) < C_i(v)$ and $C_j(w) \leq C_j(v)$ for all $j \leq i$. Note that $\prec_{\mathsf{CIRC}}$ is a prioritized version of the preference relation from *cardinality-based* circumscription (Liberatore and Schaerf 1997; Moinard 2000). The associated preorder $\preceq_{\mathsf{CIRC}}$ can be seen to satisfy the system of spheres definition.

Definition 6 ($\Vdash\!\sim$)**.** For w a world, b a time, and ϕ an $\mathcal{LL}$ sentence, we define $\Vdash\!\sim$ by $w, b \Vdash\!\sim \phi$ if $\preceq_{\mathsf{CIRC}}, w, b \Vdash\!- \phi$, and we define $\Vdash\!\sim \phi$ if $\preceq_{\mathsf{CIRC}}, w \Vdash\!- \phi$ for every world w.

Using the fixed system of spheres $\preceq_{\mathsf{CIRC}}$, essentially the reader represented using the $\mathbf{B}_\psi$ operator "only knows" the knowledge base ψ (see Levesque (1990)) but also applies circumscription to determine their beliefs. So we have, e.g., $\Vdash\!\sim \mathbf{B}_{(P \supset ab)} \neg P$. Note that Observations 2 and 3 still apply if the use of '$\Vdash\!-$' in them is replaced by '$\Vdash\!\sim$', and Observation 1 works for any system of spheres, including $\preceq_{\mathsf{CIRC}}$.

4 Examples of formalizing reader knowledge

As a reader reads, they draw conclusions about the imaginary world of the story, and also about real-world literary truths, like the genre of a story. Consider the knowledge base $\psi = \mathbf{I}(\forall x(\mathit{Knight}(x) \wedge \neg Ab(x) \supset \mathit{Man}(x))) \wedge (\mathbf{I}(\exists x \mathit{Dragon}(x)) \supset \mathit{FantasyGenre})$ which states that (in imagination) knights are normally men, and that the existence of (imaginary) dragons is a sign of the story belonging to the fantasy genre. If a reader with this knowledge reads as the first sentence of discourse $(\mathit{Dragon}(^{\#}1) \wedge \mathit{Knight}(^{\#}2))$, which is a formal version of "There was a knight and a dragon", we would want them to believe that the story is a fantasy and that there is in imagination a man, and that is what we have: $\Vdash\!\sim [\mathit{Dragon}(^{\#}1) \wedge \mathit{Knight}(^{\#}2)](\mathbf{B}_\psi \mathit{FantasyGenre} \wedge \mathbf{B}_\psi \mathbf{I} \mathit{Man}(^{\#}2))$.

Expectations about the story's development can also be represented. The expectation that a story will literally follow a version of the rule of "Chekhov's gun" – if a gun is shown hanging on the wall in one scene, it should be fired by the end of the story – can be written as $\forall x \forall e_1 \exists e_2 \big(\mathit{Mentioned}(e_1) \wedge \mathbf{I}(\mathit{HangingOnWall}(x, e_1) \wedge \mathit{Gun}(x)) \wedge \neg Ab \supset (\mathit{Mentioned}(e_2) \wedge \mathbf{I}(\mathit{Firing}(x, e_2)))\big)$. That is, if the eventuality e_1 of a gun hanging on the wall is mentioned, then normally a firing event e_2 is also mentioned. (The reader would also need further knowledge about events to prevent considering that $e_1 = e_2$.) How to encode the general underlying pragmatic principle is less clear.

4.1 Carry-over and genre conventions

In the example with the knight, to make the belief that knights are normally men applicable to fiction, we enclosed it in an $\mathbf{I}$ operator. However, we would prefer to write beliefs about the real world, and have them automatically get carried over to fiction. A first, syntactic, approximation to that is the following: Suppose the knowledge base ψ is a conjunction including a conjunct $\forall \vec{x}(\phi(\vec{x}))$ ($\vec{x}$ abbreviates the sequence of all leading universally quantified variables). If $\phi(\vec{x})$ uses only operators from first-order logic and does not include real atoms for which there are not imaginary counterparts, then $\mathbf{I}(\phi(\vec{x}))$ is also a formula. Then you could automatically generate the sentence $\forall \vec{x}(Ab(\vec{x}) \vee \mathbf{I}(\phi(\vec{x})))$, where Ab is some abnormality predicate (of appropriate arity) not used in ψ. This new sentence, roughly a defeasible imaginary copy of $\forall \vec{x}(\phi(\vec{x}))$, could be conjoined with ψ.

Carry-over by humans is probably more complicated than that. Ryan (1991, ch. 3) proposed restrictions on what should get carried over, including that the existence of real people or geographic locations should only be carried over into fictions that name at least one real person or location. A psychological experiment of Weisberg and Goldstein (2009) suggested that people are less likely to carry over facts into fictions differing from reality in other ways.

Below we consider the interaction of carry-over with fictional conventions in two examples of philosophical origin.

Scrulch the dragon Lewis (1978, p. 45) gave a case where fictional truth depends on more than world knowledge:

> Suppose I write a story about the dragon Scrulch, a beautiful princess, a bold knight, and what not. It is a perfectly typical instance of its stylized genre, except that I never say that Scrulch breathes fire. Does he nevertheless breathe fire in my story? Perhaps so, because dragons in that sort of story do breathe fire. But the explicit content does not make him breathe fire. Neither does background, since in actuality and according to our beliefs there are no animals that breathe fire.

For us there is no difficulty in writing additional sentences that describe how things in imagination are different from in reality, such as $Ab \vee \mathbf{I}(\forall x(\mathit{Dragon}(x) \supset \mathit{BreathesFire}(x)))$. Here Ab would represent the abnormality of a story about dragons which didn't breathe fire. We would have to give Ab sufficiently high priority so that this sentence would overrule any carried over beliefs about animals not breathing fire in general. Note that the sentence does not do anything to specify the fire-breathing abilities of real dragons; despite believing that fictional dragons normally breathe fire, the reader could still regard real dragons that breathe fire as (even) less plausible than real dragons that do not breathe fire.

Recognizing a witch Walton (1990, §4.3) gave a number of examples of tricky cases about fictional truth, including one about what information is needed to recognize a fictional character as a witch. He wrote (p. 161, 164) the following (about drawing, but clearly also relevant to other media):

> Any child can draw a witch. Depicting a woman with a black cape, conical hat, and long nose will usually do the trick. [...] The fact that fictionally there is a witch is implied by the fact that fictionally there is a woman with a black cape, conical hat, and long nose. But it is not the case that were there (in the real world) a long-nosed woman decked out in black cape and conical hat [...], there would be a witch. [...] Although it is fictional in a mutually recognized legend that there are witches and that they have long noses and wear conical hats, it is much less clearly fictional in it that, were there a

> woman of that description, she would be a witch. Is it part of the legend that there are no Halloween parties, or that nonwitches never dress thus [...]?

This idea, that someone described in a stereotypically witch-like way is a witch while not necessarily everyone in the world of the story with witch-like characteristics is a witch, can be expressed in literary logic. We can do so by writing $\forall x(\mathbf{I}(\mathit{WitchLike}(x)) \wedge \mathit{Mentioned}(x) \wedge \neg \mathit{Ab}(x) \supset \mathbf{I}\,\mathit{Witch}(x))$. Then the $\mathbf{I}\,\mathit{Witch}(x)$ conclusion is not drawn for every x having witch-like characteristics, but only for those also mentioned in the discourse (recall the special $\mathit{Mentioned}$ predicate). This is an example of *scoped* non-monotonic reasoning (Etherington, Kraus, and Perlis 1991).

4.2 Expectations about mystery stories

A reader may expect when reading a mystery story to eventually find out who is guilty. In this section, we will use the unary imaginary predicate symbol $G(x)$ to mean that x is guilty (in imagination).

Suppose that ψ is the KB of a reader, and we want to inform this reader that they should expect to find out who is guilty. The proposition below shows how we can extend ψ into a KB ψ' so a reader knowing only ψ' believes they will find out who is guilty (assuming that a reader knowing only the original KB, ψ, does *not* believe that they *won't* find out who's guilty – in other words, that $\Vert\!\!\sim \neg\mathbf{B}_\psi\neg\exists x(\mathbf{F}_\psi\, G(x))$).

Proposition 1. Let G be a unary imaginary predicate. Suppose ψ is an $\mathcal{LL}$ sentence s.t. $\Vert\!\!\sim \neg\mathbf{B}_\psi\neg\exists x(\mathbf{F}_\psi\, G(x))$. Let $\psi' = \psi \wedge \exists x(\mathbf{F}_\psi\, G(x))$. Then $\Vert\!\!\sim \mathbf{B}_{\psi'}\exists x(\mathbf{F}_{\psi'}\, G(x))$.

Proof. We want to prove that for every world w, we have $w, 0 \Vert\!\!\sim \mathbf{B}_{\psi'}\exists x(\mathbf{F}_{\psi'}\, G(x))$. To do that, we want to show that $v, 0 \Vert\!\!\sim \exists x(\mathbf{F}_{\psi'}\, G(x))$ for every $v \in \min([w]{\sim}^{\psi'}_0)$. Fix an arbitrary such v (if there are none, we are done), and let $n = |v_d|$. We have that $v, 0 \Vert\!\!\sim \psi'$ and so (by the definition of ψ') $v, 0 \Vert\!\!\sim \psi$ and $v, 0 \Vert\!\!\sim \exists x(\mathbf{F}_\psi\, G(x))$. Let $c \in \mathcal{N}$ be such that $v, 0 \Vert\!\!\sim \mathbf{F}_\psi\, G(c)$. Then $v, n \Vert\!\!\sim \mathbf{B}_\psi \mathbf{I} G(c)$. Therefore, for each $v' \in \min([v]{\sim}^{\psi}_n)$, we have $v', n \Vert\!\!\sim \mathbf{I} G(c)$. It can be shown that $\min([v]{\sim}^{\psi'}_n) \subseteq \min([v]{\sim}^{\psi}_n)$, which means that for each $v^* \in \min([v]{\sim}^{\psi'}_n)$, we have $v^*, n \Vert\!\!\sim \mathbf{I} G(c)$. Hence $v, n \Vert\!\!\sim \mathbf{B}_{\psi'}\mathbf{I}G(c)$, and $v, 0 \Vert\!\!\sim \exists x(\mathbf{F}_{\psi'}\, G(x))$. $\square$

We could also consider representing the knowledge that the reader won't find out who's guilty until very near the end, which Brewer and Lichtenstein (1982) considered to be an example of a "curiosity discourse organization". The non-trivial part would be formalizing the vague "very near" (to say the reader won't have a belief about who's guilty at a precise time – say, three sentences before the end – we could simply write something like $\blacktriangleright\!\!\mid \bullet\bullet\bullet\neg\exists x \mathbf{B}_\psi \mathbf{I} G(x)$). Brewer and Lichtenstein suggested that the purpose of stories is to entertain, and that three ways that authors accomplish this is by creating suspense, surprise, and curiosity by manipulating when information gets revealed to the reader. Our next example might be considered a case of surprise.

A genre-savvy reader might think that in a mystery story, it's true that "The first character the author tries to make you suspect of being guilty is innocent." Under an assumption of authorial competence we can roughly paraphrase that as "The first character a naïve reader would suspect of being guilty is innocent." Consider a reader with KB ψ; let us suppose that they are the reader the author would be able to trick into suspecting the wrong character. How could we extend their KB to make them genre-savvy?

As a prelude to that, consider the following formula:

$$\phi(x) = \mathbf{B}_\psi \mathbf{I} G(x) \wedge \forall y(y \neq x \supset \neg\blacklozenge\mathbf{B}_\psi\mathbf{I}G(y)) \tag{4}$$

Recalling that "$\blacklozenge$" is the "previously" operator, we can read $\phi(x)$ as "x is believed to be guilty (in imagination) and for all y not equal to x, y was not previously believed to be guilty (in imagination)", where the beliefs are understood to be those of the reader with KB ψ. So, for a standard name c, $\phi(c)$ is true at a time iff c is the (unique) first character believed to be guilty (in imagination).

Below, proposition 2 shows a sentence (incorporating $\phi(x)$ as a subformula) we can conjoin to ψ to produce the knowledge base ψ' of a savvy reader, and establishes that if $\phi(c)$ is ever true (i.e., that c is the first character the reader with KB ψ believes is guilty) then the reader with KB ψ' will believe that c is not guilty (unless that reader *knows* that c is guilty, e.g. because the discourse includes $G(c)$ explicitly).

Proposition 2. Let ψ be an $\mathcal{LL}$ sentence, let Ab be a 0-ary real abnormality predicate of higher priority than any abnormality predicate appearing in ψ, let G be a unary imaginary predicate, and (as in Equation 4) let $\phi(x) = \mathbf{B}_\psi\mathbf{I}G(x) \wedge \forall y(y \neq x \supset \neg\blacklozenge\mathbf{B}_\psi\mathbf{I}G(y))$. Then define $\psi' = \psi \wedge \big(Ab \vee \Box\forall x(\phi(x) \supset \mathbf{I}\neg G(x))\big)$. Then

$$\Vert\!\!\sim \Box\forall x\big((\phi(x) \wedge \neg\mathbf{K}_{\psi'}\mathbf{I}G(x)) \supset \mathbf{B}_{\psi'}\mathbf{I}\neg G(x)\big).$$

Proof. Suppose for a world w, time b, and name c, we have $w, b \Vert\!\!\sim \phi(c) \wedge \neg\mathbf{K}_{\psi'}\mathbf{I}G(c)$. We want to show that $w, b \Vert\!\!\sim \mathbf{B}_{\psi'}\mathbf{I}\neg G(c)$. It can be shown that $w, b \Vert\!\!\sim \neg\mathbf{K}_{\psi'}Ab$, and therefore (because all worlds in which Ab is false are more plausible than all others) $w, b \Vert\!\!\sim \mathbf{B}_{\psi'}\neg Ab$. So $w, b \Vert\!\!\sim \mathbf{B}_{\psi'}\Box\forall x(\phi(x) \supset \mathbf{I}\neg G(x))$, and so $w, b \Vert\!\!\sim \mathbf{B}_{\psi'}(\phi(c) \supset \mathbf{I}\neg G(c))$. So we will be done if we can show that $w, b \Vert\!\!\sim \mathbf{B}_{\psi'}\phi(c)$, i.e. that $v, b \Vert\!\!\sim \phi(c)$ for every $v \in \min([w]{\sim}^{\psi'}_b)$. This follows because for any such v, the discourse v_d must agree with w_d on the first b entries, which is enough to give $\phi(c)$ the same truth value there. $\square$

5 Discussion and related work

5.1 Regarding non-monotonic reasoning

We have space only to make a couple remarks in this section.

Many forms of circumscription, including prioritized circumscription, have been considered in the literature (see e.g. (Lifschitz 1994)). Cardinality-based circumscription (Liberatore and Schaerf 1997; Moinard 2000) has the advantage of being simple to work with because there will always be a set of most plausible worlds in which any sentence is true, if that sentence is true in any worlds.

The way that we can use the ψ in the $\mathbf{B}_\psi$ operator to refer to what is and isn't believed by an agent with knowledge ψ' recalls hierarchic autoepistemic logic (Konolige 1988). In common with that system, and unlike standard autoepistemic logic (Levesque 1990), the issue of there being multiple "stable expansions" of an agent's beliefs does not arise.

5.2 Regarding stories and fiction

Wilensky (1983) briefly discussed "dynamic points" in stories, involving violations of the expectations of a character or the reader. He wrote (p. 616) that "Only with recourse to events that are supposed to transpire in the reader during the course of understanding a text can the discourse structure of a theory of stories be stated." $\mathcal{LL}$, of course, is expressly designed to allow referring to changing beliefs of the reader.

Michael (2013) also considered encoding reader expectations in a formal system. He gave an example of encoding the expectation "that the story clarifies at each instance whether it is day or night" (which is not unreasonable for a story told in a visual medium), though the brief outline given of the semantics for his system does not cover how disjunctive expectations like that should be handled. Michael also considers the case of the reader being told additional information about what the author expects them to infer.

We may note that there is also work in the AI subfield of narrative *generation* that concerns itself with reader expectations. For example, the "Prevoyant" system (Bae and Young 2014) is supposed to generate narratives that are surprising.

Rapaport and Shapiro (1995) presented a computational approach to handling carry-over in story understanding in their SNePS system. Beliefs about reality are copied into a "story world context" and belief revision is used to deal with any conflicts that may arise as the story is read (they also discuss an alternative approach using a "story operator"). Genre conventions are not discussed, and since time for the reader does not play an explicit role, it's not clear how reader beliefs about the future could be represented or queried.

The ISAAC story understanding system (Moorman and Ram 1994) has a so-called "creative understanding" process that allows for modifying pre-existing concepts in an attempt to understand a story. Moorman and Ram did not relate this to the philosophical literature on carry-over, and further investigation of that would be interesting.

Bonomi and Zucchi (2003) gave an approach to combining carry-over with genre conventions. They consider having two systems of spheres (the worlds in these, unlike our complex worlds, are not split into real and imaginary parts), one centered on B_x, the set of worlds conforming to the "overt beliefs" of the author of x (the fiction in question), and one centered on R_x, the set of worlds following the conventions for x. Fictional truth is determined by what is true in all the closest worlds to B_x from among those worlds that are closest to R_x in which the "directly generated content"[3] of x is true. Note this requires which conventions x follows to be already known, while $\mathcal{LL}$ allows for reasoning about that.

We have not much considered interaction between expectations and carry-over. However, Martínez-Bonati (1983) suggested that the reader expects to quickly find out how realistic the world of the story is (p. 188):

> If I read a few narrative sentences implying a system of reality not different from ordinary life, I will rapidly tend to solidify my expectations into a "realistic" fictional horizon. [...] A similar promptness will be an attribute of the projection of fictional horizons that are traditional and well-known (for example, the fabulous world of speaking animals).

At an intermediate point in reading, the reader's beliefs about what carries over may be influenced not just by what the author has said about the fictional world, but by what the reader believes the author *will* say in the rest of the story.

Our formal discourses require temporal information be explicitly encoded, like some other approaches (Diakidoy et al. 2014; Maslan, Roemmele, and Gordon 2015). While this is not like the ordinary use of natural language, it is much less complicated. For logic-based approaches that do try to deal with those kinds of issues, see Episodic Logic (Schubert and Hwang 2000) and Segmented Discourse Representation Theory (Asher and Lascarides 2005).

The form of fictional truth we have formalized is reader-dependent. Whether what a text means should depend on the reader is controversial (Hirst 2008). An objective form might be implemented in a multi-agent version of literary logic, by defining objective fictional truth in terms of the knowledge of an ideal reader which other readers have beliefs about. ($\mathcal{LL}$ in this paper is not truly multi-agent, despite the parameterized $\mathbf{B}_\psi$ operators, since one reader cannot reason about another's beliefs without specifying what the latter's KB is.)

[3] This is not the same as the literal content, as they also consider (in an unformalized way) the narrator and their reliability.

6 Conclusion

We have argued that our logic $\mathcal{LL}$, following on the work of Friedman and Halpern, can represent various forms of knowledge relevant to story understanding, and so be used to determine how a reader with such knowledge would answer questions about a story. We encourage investigating how other formal approaches developed for modelling belief over time could similarly be useful in story understanding.

In future work, we plan to further investigate carry-over and to construct fully worked out examples with complete stories. Also, it should be possible to replace the $Mentioned$ predicate with epistemic constructions. Another point is that $\mathcal{LL}$ models the reader as *logically omniscient* (Hintikka 1975), seeing all consequences of its own beliefs, but it would be interesting to consider resource-bounded readers, as that is what real authors write for.

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

Asher, N., and Lascarides, A. 2005. *Logics of Conversation*. Studies in natural language processing. Cambridge University Press.

Bae, B.-C., and Young, R. M. 2014. A computational model of narrative generation for surprise arousal. *IEEE Transactions on Computational Intelligence and AI in Games* 6(2):131–143.

Bonomi, A., and Zucchi, S. 2003. A pragmatic framework for truth in fiction. *Dialectica* 57(2):103–120.

Brewer, W. F., and Lichtenstein, E. H. 1982. Stories are to entertain: A structural-affect theory of stories. *Journal of Pragmatics* 6(56):473–486.

Chambers, N., and Jurafsky, D. 2009. Unsupervised learning of narrative schemas and their participants. In *ACL 2009*, 602–610.

Charniak, E., and Goldman, R. 1989. Plan recognition in stories and in life. In *UAI 1989*, 54–59.

Charniak, E. 1972. Toward a model of children's story comprehension. MIT AI Laboratory Technical Report 266.

Chaturvedi, S.; Peng, H.; and Roth, D. 2017. Story comprehension for predicting what happens next. In *EMNLP 2017*, 1604–1615.

Diakidoy, I.-A.; Kakas, A.; Michael, L.; and Miller, R. 2014. Story comprehension through argumentation. In *Computational Models of Argument - Proceedings of COMMA 2014*, 31–42.

Doležel, L. 1995. Fictional worlds: Density, gaps, and inference. *Style* 29(2):201–214.

Etherington, D. W.; Kraus, S.; and Perlis, D. 1991. Nonmonotonicity and the scope of reasoning. *Artificial Intelligence* 52(3):221–261.

Friedman, N., and Halpern, J. Y. 1999. Modeling belief in dynamic systems, part II: Revision and update. *Journal of Artificial Intelligence Research (JAIR)* 10:117–167.

Grove, A. 1988. Two modellings for theory change. *Journal of Philosophical Logic* 17(2):157–170.

Heintz, J. 1979. Reference and inference in fiction. *Poetics* 8(1):85–99.

Hintikka, J. 1975. Impossible possible worlds vindicated. *Journal of Philosophical Logic* 4(4):475–484.

Hirst, G. 2008. The future of text-meaning in computational linguistics. In *Proceedings of the 11th International Conference on Text, Speech and Dialogue (TSD 2008)*, 3–11.

Hobbs, J. R.; Stickel, M.; and Martin, P. 1993. Interpretation as abduction. *Artificial Intelligence* 63:69–142.

Hobbs, J. R. 1990. *Literature and Cognition*. Number 21 in CSLI Lecture Notes. Center for the Study of Language and Information.

Iyyer, M.; Manjunatha, V.; Guha, A.; Vyas, Y.; Boyd-Graber, J.; Daumé III, H.; and Davis, L. 2017. The amazing mysteries of the gutter: Drawing inferences between panels in comic book narratives. In *CVPR 2017*.

Konolige, K. 1988. Hierarchic autoepistemic theories for nonmonotonic reasoning. In *AAAI 1988*, 439–443.

Levesque, H. J., and Lakemeyer, G. 2000. *The logic of knowledge bases*. MIT Press.

Levesque, H. J. 1990. All I know: A study in autoepistemic logic. *Artificial Intelligence* 42(2-3):263–309.

Lewis, D. 1973. *Counterfactuals*. Harvard University Press.

Lewis, D. 1978. Truth in fiction. *American Philosophical Quarterly* 15(1):37–46.

Liberatore, P., and Schaerf, M. 1997. Reducing belief revision to circumscription (and vice versa). *Artificial Intelligence* 93(1):261–296.

Lichtenstein, O.; Pnueli, A.; and Zuck, L. D. 1985. The glory of the past. In *Proceedings of the Conference on Logic of Programs*, 196–218.

Lifschitz, V. 1994. Circumscription. In *Handbook of Logic in AI and Logic Programming*, volume 3. Oxford University Press. 298–352.

Martínez-Bonati, F. 1983. Towards a formal ontology of fictional worlds. *Philosophy and Literature* 7(2):182–195.

Maslan, N.; Roemmele, M.; and Gordon, A. S. 2015. One hundred challenge problems for logical formalizations of commonsense psychology. In *Commonsense 2015*, 107–113.

McCarthy, J. 1986. Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence* 28(1):89–116.

McCarthy, J. 1990. An example for natural language understanding and the AI problems it raises. In Lifschitz, V., ed., *Formalizing Common Sense: Papers by John McCarthy*. Ablex Publishing Corporation. 70–76.

Michael, L. 2013. Story understanding... calculemus! In *Commonsense 2013*.

Moinard, Y. 2000. Note about cardinality-based circumscription. *Artificial Intelligence* 119(1):259 – 273.

Moorman, K., and Ram, A. 1994. A model of creative understanding. In *AAAI 1994*, 74–79.

Morgenstern, L. 2014. Representing and reasoning about time travel narratives: Foundational concepts. In *KR 2014*, 642–645.

Mostafazadeh, N.; Chambers, N.; He, X.; Parikh, D.; Batra, D.; Vanderwende, L.; Kohli, P.; and Allen, J. 2016. A corpus and cloze evaluation for deeper understanding of commonsense stories. In *NAACL HLT 2016*, 839–849.

Quine, W. V. 1968. Propositional objects. *Crítica: Revista Hispanoamericana de Filosofía* 2(5):3–29.

Rapaport, W. J., and Shapiro, S. C. 1995. Cognition and fiction. In Duchan, J. F.; Bruder, G. A.; and Hewitt, L. E., eds., *Deixis in Narrative: A Cognitive Science Perspective*. Lawrence Erlbaum Associates, Inc. 107–128.

Richardson, M.; Burges, C. J.; and Renshaw, E. 2013. MCTest: A challenge dataset for the open-domain machine comprehension of text. In *EMNLP 2013*, 193–203.

Rumelhart, D. 1975. Notes on a schema for stories. In Bobrow, D. G., and Collins, A. M., eds., *Representation and Understanding: Studies in Cognitive Science*. Academic Press, Inc. 211–236.

Ryan, M.-L. 1991. *Possible Worlds, Artificial Intelligence, and Narrative Theory*. Indiana University Press.

Schank, R. C., and Abelson, R. P. 1977. *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Lawrence Erlbaum Associates, Inc.

Schubert, L. K., and Hwang, C. H. 2000. Episodic Logic meets Little Red Riding Hood: A comprehensive, natural representation for language understanding. In Iwanska, L., and Shapiro, S. C., eds., *Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language*. MIT/AAAI Press. 111–174.

Shapiro, S.; Pagnucco, M.; Lespérance, Y.; and Levesque, H. J. 2011. Iterated belief change in the situation calculus. *Artificial Intelligence* 175(1):165–192.

Tapaswi, M.; Zhu, Y.; Stiefelhagen, R.; Torralba, A.; Urtasun, R.; and Fidler, S. 2016. MovieQA: Understanding stories in movies through question-answering. In *CVPR 2016*.

Van Inwagen, P. 1977. Creatures of fiction. *American Philosophical Quarterly* 14(4):299–308.

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

Weisberg, D. S., and Goldstein, J. 2009. What belongs in a fictional world? *Journal of Cognition and Culture* 9(1):69–78.

Wilensky, R. 1983. Story grammars versus story points. *Behavioral and Brain Sciences* 6(4):579–623.

Wolterstorff, N. 1976. Worlds of works of art. *The Journal of Aesthetics and Art Criticism* 35(2):121–132.

Woods, J. 1974. *The Logic of Fiction: A Philosophical Sounding of Deviant Logic*. Mouton.