COGNITION AND FICTION: AN INTRODUCTION

William J. Rapaport and Stuart C. Shapiro

Department of Computer Science and Center for Cognitive Science
State University of New York at Buffalo, Buffalo, NY 14260
{rapaport | shapiro}@cs.buffalo.edu
http://www.cs.buffalo.edu/~snwiz

October 9, 1997

Abstract

1 COMPUTATIONAL PHILOSOPHY OF FICTION.

This is an essay in computational philosophy: the investigation of philosophical issues using computational methods as well as the application of philosophy to problems in computer science. The philosophical issues we will explore include predication and fiction. The computational issues are primarily in artificial intelligence (AI).

“Knowledge” representation is the study of the representation of information in an AI system; since the information need not be true—especially if the information is from fiction—a more accurate name would be “belief” representation (cf. Rapaport 1992). In “An Introduction to a Computational Reader of Narratives” (Shapiro & Rapaport 1995)—a companion piece to this chapter—we looked at how predication is represented in a “knowledge”-representation system when it is used for cognitive modeling and natural-language competence (by which we mean both natural-language understanding and generation; cf. Shapiro & Rapaport 1991). The present chapter discusses appropriate means of representing fictional items, fictional predication, and propositions from fiction in such a system. We briefly survey four philosophical ontological theories of fiction and sketch an epistemological theory of fiction using a story operator and rules for allowing propositions to “migrate” into and out of story “spaces”.

2 FICTIONAL PREDICATION.

Our colleagues and we have been investigating how a cognitive agent is able to read a narrative and comprehend the indexical information in it: where the events described in the narrative are taking place (i.e., where in the “story world”—a semantic domain corresponding to the syntactic narrative text), when
they take place (in the time-line of the story world), who the participants in these events are (the characters in the story world), and from whose point of view the events and characters are described (see Duchan et al. 1995).

In order to do this, a reader (human or machine) has to be able to (1) read a narrative (in particular, a fictional narrative), (2) construct a mental representation or model of the story and the story world, and (3) use that mental model to understand and to answer questions about the story. To construct the mental model, she will need to contribute something to her understanding of the narrative. One contribution is in the form of the “deictic center”—a data structure that contains the indexical information needed to track the who, when, and where.

Another contribution is background knowledge about the real world. For instance, in reading a novel about the Civil War, a reader would presumably bring to her understanding of it some knowledge of the Civil War, such as that Abraham Lincoln was the 16th president of the U.S. and was assassinated in 1865, even if that information is not explicitly stated in the novel. The novel might go on to make other claims about Lincoln, such as that he was tall or that he had a particular conversation with General Grant on a particular day in 1860 (even if, in fact, they never talked on that day—this is a novel, after all). Such a claim would probably not be inconsistent with anything the reader antecedently believed about Lincoln. But some claims in the novel might be inconsistent in this way, e.g., if she read that Lincoln was re-elected to a third term in 1868. So the reader has to be able to represent the information presented in the narrative, keep it suitably segregated from her background knowledge, yet be able to have information from her antecedent real-world beliefs “migrate” into her model of the story world as well as have information from the story world “migrate” back into her store of beliefs about the real world: There must be a semi-permeable membrane separating these two subspaces of her mental model (cf. Yordy 1990–1991: 2).

There have been a number of theories in philosophy about the nature of fictional objects. All of these are ontological theories concerned with such questions as: What are fictional objects? How are properties predicated of them? How are fictional objects related to non-fictional ones? However, for the purposes of our project, we need to be more concerned with “epistemological” or processing/computational/interpretive issues: How does a reader understand a (fictional) narrative? How does a reader decide whether and to what extent it is fictional? How does a reader construct a mental model of the story world? How does a reader represent fictional entities and the properties predicated of them? How do readers integrate their knowledge of the real world with what they read in the narrative? And so on. Some of these are, indeed, ontological issues, but they are what we have elsewhere termed issues in “epistemological ontology” (Rapaport 1985/1986): Corresponding to the purely or metaphysically ontological question, “What are fictional objects?”, we ask the epistemologically ontological question, “How does a cognitive agent represent fictional objects?”. And corresponding to the purely ontological question, “How are properties predicated of fictional objects?”, we ask the epistemologically ontological question, “How does a cognitive agent represent the predication of properties of fictional objects?”.

In this chapter, we examine several philosophical theories of fiction to see what aspects of them are useful for our cognitive/computational project, and we propose a representation scheme that satisfactorily answers most of the kinds of questions raised above (and that incorporates an exciting, if counterintuitive, proposal for the remaining questions). Our scheme is implemented in SNePS, an intensional, propositional, semantic-network “knowledge”-representation and reasoning system that is used for research in AI and in cognitive science. (The uninformed reader will benefit from reading the companion article, Shapiro & Rapaport 1995.) Specifically, the proposed representation scheme is to embed the propositions of the fictional narrative in a “story operator” that is formally akin to the belief representations we already have in SNePS (Rapaport 1986; Wiebe & Rapaport 1986; Rapaport, Shapiro, & Wiebe 1997). We will show how SNePS’s propositional and fully intensional nature, plus the story operator, allow the best aspects of the philosophical theories to be implemented.
2.1 Four Ontological Theories of Fiction.

Let us begin by briefly surveying four (out of many, many more) philosophical theories of the ontological status of fictional objects. We will not be concerned so much with criticizing them (though we will mention some difficulties they have), as we will with finding what aspects of them might be useful for our, rather different, purposes.

2.1.1 Castañeda’s theory.

Hector-Neri Castañeda’s theory of guises and consubstantiation is an all-encompassing theory of the objects of thought as well as of the objects in the world (Castañeda 1972, 1975ab, 1977, 1980, 1989); it includes a theory of fictional objects (Castañeda 1979, 1989). We have discussed the full theory in more detail elsewhere (Rapaport 1978, 1985a), so we will content ourselves with a presentation of his theory of fiction here.

Castañeda takes a uniform viewpoint, with which we agree: All objects in fiction are to be treated alike, whether they are “real” or “fictional” (cf. Schöles 1968, Rapaport 1985a). They are, in his terminology, “guises”, i.e., roughly, intensional objects of thought. But there are different modes of predication of properties to guises. Thus, if one reads a narrative about the Civil War that Lincoln died in 1865, this would be analyzed in his theory as a “consubstantiation” (C*) of two guises, the guise c{being Lincoln} (i.e., the intensional object of thought whose sole internal property is being Lincoln) and the guise c{being Lincoln, having died in 1865} (i.e., the intensional object of thought whose sole internal properties are being Lincoln and having died in 1865):

\[ C^*(c\{\text{being Lincoln}\}, c\{\text{being Lincoln, having died in 1865}\}) \]

Consustantiation is an existence-entailing equivalence relation. On the other hand, if one reads another narrative, in which the author has stated that Lincoln was re-elected in 1868, this would be analyzed as a “consociation” (C**) of two guises:

\[ C^{**}(c\{\text{being Lincoln}\}, c\{\text{being Lincoln, having been re-elected in 1868}\}) \]

Consociation is an equivalence relation, not entailing existence, among guises that are “joined together” in a mind. But it is the same Lincoln (i.e., c{being Lincoln}) in both cases.

That is a rather drastic oversimplification, but it raises the following concern: How is the reader to decide whether a sentence read in the course of a narrative is to be analyzed by consubstantiation or by consociation? In fact, we would claim, the uniformity with respect to the objects should be extended to the mode of predication: All predications in narrative are consociational, even the “true” ones.

Castañeda also admits the existence of “story operators” into his theory, but finds them otiose. A story operator is a (usually modal) operator that prefixes all sentences in a narrative: “In story S, it is the case that \( \varphi \). Not all theorists of fiction find them attractive (cf. Rapaport 1976, 1985a), but, as Castañeda points out, one can hardly deny that they exist: One can take the operator to be the title page of the narrative! His claim is that story operators fail to account for the interesting or problematic aspects of fiction.

An example in the context of SNePS might clarify things a bit. Consider the situation illustrated in Figure 1. Suppose that the reader has a background belief (“world knowledge”, as we might say) that

(1) George Washington was the first president.

This would be analyzed as a consubstantiation. Suppose next that the cognitive agent reads in a narrative that

\[ \text{(1) George Washington was the first president.} \]
(2) George Washington chopped down a cherry tree.

This would be analyzed as a consociation. The processing problem is this: If both sentences were to have occurred in the narrative, they would have to be treated alike, using the same mode of predication, namely, consociation. But this is a reasonable modification of Castañeda’s theory, and there are no other problems so far, so all is well.

2.1.2 Lewis’s theory.

David Lewis’s theory of fiction (Lewis 1978) makes essential use of the story operator, and, despite earlier misgivings about them (see the references above), we now find that they have a useful role to play. But Lewis’s version has some problems. He allows his story operator to be dropped by way of abbreviation. Thus, we might say

Sherlock Holmes lived at 221B Baker Street,

but what we really mean is, e.g.,

In *The Hound of the Baskervilles*, Sherlock Holmes lived at 221B Baker Street,

since, after all, the former is false and the latter is true.

There is an evident advantage to this, for it enables us to distinguish between “facts” about fictional and non-fictional entities: a worthy endeavor, and one that the reader must be able to do. In fact, she will do it much the way that Lewis recommends. Consider the following argument:

\[
\begin{align*}
\text{Lived-at}(221B \text{ Baker St.}, \text{ Sherlock Holmes}) \\
221B \text{ Baker St.} = \text{ a bank} \\
\therefore \text{ Lived-at}(\text{ a bank}, \text{ Sherlock Holmes})
\end{align*}
\]

Although the first premise is true in the story world (but false or truth-valueless in the real world), and the second is factually true (cf. Rule 1989), the conclusion is false in both the real world and the story world. But merely replacing the story operator won’t help:

\[
\begin{align*}
\text{In } \text{The Hound of the Baskervilles}, \text{ Lived-at}(221B \text{ Baker St.}, \text{ Sherlock Holmes}) \\
221B \text{ Baker St.} = \text{ a bank} \\
\therefore \text{ In } \text{The Hound of the Baskervilles}, \text{ Lived-at}(\text{ a bank}, \text{ Sherlock Holmes})
\end{align*}
\]

fares no better, since 221B Baker St. is not a bank in *The Hound of the Baskervilles*. Nor does:

\[
\begin{align*}
\text{In } \text{The Hound of the Baskervilles}, \text{ Lived-at}(221B \text{ Baker St.}, \text{ Sherlock Holmes}) \\
221B \text{ Baker St.} = \text{ a bank} \\
\therefore \text{ In } \text{The Hound of the Baskervilles}, \text{ Lived-at}(\text{ a bank}, \text{ Sherlock Holmes})
\end{align*}
\]

fare any better, since the conclusion is false with or without the story operator. But a uniform application of the story operator works fine:

\[
\begin{align*}
\text{In } \text{The Hound of the Baskervilles}, \text{ Lived-at}(221B \text{ Baker St.}, \text{ Sherlock Holmes}) \\
\text{In } \text{The Hound of the Baskervilles}, 221B \text{ Baker St.} = \text{ a bank} \\
\therefore \text{ In } \text{The Hound of the Baskervilles}, \text{ Lived-at}(\text{ a bank}, \text{ Sherlock Holmes})
\end{align*}
\]
and

Lived-at(221B Baker St., Sherlock Holmes)
221B Baker St. = a bank
\therefore Lived-at(a bank, Sherlock Holmes)

are both valid, albeit unsound: The former is unsound, because the second premise is false; the latter is unsound, because the first premise is false.

The difficulty with Lewis's proposal is that

Sherlock Holmes is fictional

is false no matter how you slice it. It's false with the story operator restored, since, within the story, Holmes is as real as anyone is. And it's false (or at least truth-valueless) without it, since 'Sherlock Holmes' is a non-denoting expression. This difficulty is unacceptable.

2.1.3 Parsons's theory.

Terence Parsons's theory of fiction (Parsons 1975, 1980) is based on his theory of nonexistent objects. In contrast to Castañeda, whose theory has one kind of property but two modes of predication, Parsons's has two kinds of properties (nuclear and extranuclear) but only one mode of predication. Rather than rehearse his full theory of fiction here (see Rapaport 1985a for a summary and critique), we shall focus on a distinction he makes between "native", "immigrant", and "surrogate" fictional objects.

Native fictional objects are, roughly, those who originate in the story in which they are found, such as Sherlock Holmes in The Hound of the Baskervilles. Immigrant fictional objects are, roughly, those who have migrated into a story from elsewhere, such as London in The Hound of the Baskervilles or Sherlock Holmes in The Seven-Per-Cent Solution (Meyer 1974). But, of course, the London of The Hound of the Baskervilles has properties that the real London lacks (and vice versa), which raises obvious difficulties. So the London-of-The-Hound-of-the-Baskervilles is a surrogate fictional object, distinct from the real London.

Such distinctions can be made and are no doubt useful. But there are a number of questions that have to be answered before one can accept them: Which London did Conan Doyle discuss? Which London did Sherlock Holmes and Dr. Watson discuss? When is one discussing London and when the London-of-The-Hound-of-the-Baskervilles? In general, how does the reader distinguish properties of the "real" London from properties of the London-of-The-Hound-of-the-Baskervilles? These are questions that can be dealt with, we believe, in the SNePS proposal to be introduced below.

2.1.4 Van Inwagen's theory.

The final theory of fictional objects in our brief survey is one that we find quite congenial in many respects, though it, too, falls short. Peter van Inwagen's theory (Van Inwagen 1977), like Castañeda's, distinguishes between two modes of predication, and, like Lewis's, it uses something like a story operator.

Van Inwagen's two modes of predication are "predication" and "ascription". "Sherlock Holmes is fictional" expresses a property predicated of an existing theoretical entity of literary criticism, namely, Sherlock Holmes. (Other kinds of theoretical entities of literary criticism include novels, short stories, etc.) In contrast, "Sherlock Holmes is a detective" expresses (perhaps elliptically) a property ascribed to the same theoretical entity of literary criticism in a work of fiction:

A(detective, Sherlock Holmes, The Hound of the Baskervilles).
Note that the story is not strictly speaking a logical operator, but an essential argument place in a 3-place predication relation.

There are two problems with this otherwise quite nice theory. They are, we believe, not serious problems and could be easily gotten around. First, in “Sherlock Holmes Confronts Modern Logic” (Hintikka & Hintikka 1983), the authors call Holmes a “great detective” (p. 155). According to van Inwagen’s theory, contrary to what one might expect, it is not the case that

\[ \text{A(\text{great detective, Sherlock Holmes, “Sherlock Holmes Confronts Modern Logic”}).} \]

Why? Because “Sherlock Holmes Confronts Modern Logic” is not literature and, hence, not a theoretical entity of literary criticism. This strikes us as an unnecessary aspect of van Inwagen’s theory.

Second, assume that in *War and Peace* it is stated that Napoleon is vain. But, according to van Inwagen’s theory and again contrary to what one might expect, it is not the case that

\[ \text{A(\text{vain, Napoleon, War and Peace),}} \]

because Napoleon is not a theoretical entity of literary criticism! Again, this strikes us as unnecessary.

### 2.2 A SNePS Approach to Fiction.

In order for our computational cognitive agent (we call her “Cassie”) to read a narrative, the representations she should construct will include a story operator (as in Lewis’s or van Inwagen’s theory), only one mode of predication (as in Parsons’s theory), and only one kind of property (as in Castañeda’s theory). Since, at the time of writing, this theory is only beginning to be implemented, there is a strong possibility that this will prove insufficient: The one addition that we can foresee (urged in earlier writings, e.g., Rapaport 1976, 1985a, and suggested in conversation by Johan Lammens) is the need to distinguish between real-world entities and their surrogates; but it must be kept in mind that all entities represented in the mind of our reader are just that—entities in her mind—not entities some of which are real and some of which are fictional.

The story operator will set up a “story space” that is formally equivalent to a belief space (cf. Rapaport 1986; Wiebe & Rapaport 1986; Shapiro & Rapaport 1991, 1995; Rapaport, Shapiro, & Wiebe 1997). It will allow the reader to distinguish her own beliefs about London from (her beliefs about) claims made about London in a story in precisely the same way that belief spaces allow our computational cognitive agent to distinguish her own beliefs about John from her beliefs about Mary’s beliefs about John (as in Shapiro & Rapaport 1993, Fig. 2b; cf. Rapaport 1986; Shapiro & Rapaport 1987; Rapaport, Shapiro, & Wiebe 1997).

But how should this be handled? Consider Figure 2. Suppose that one of Cassie’s background beliefs is that Lincoln died in 1865 and that she reads in a narrative that Lincoln was re-elected in 1868. There is a processing problem: Cassie, our computational cognitive agent, is faced with an inconsistency. There are two solutions. First, the SNePS Belief Revision system (SNeBR; Martins & Shapiro 1988)—a facility for detecting and removing inconsistent beliefs—can be invoked. The detection of the inconsistency will cause a split to be made into two (consistent) contexts. But note that the net effect of this is to embed the second statement (the re-election in 1868) in a story operator. So we could start with a story operator in the first place. This is the second solution, as shown in Figure 3. (Implementations of the first solution are given in Rapaport 1991 and Rapaport & Shapiro 1995.)

But now let’s complicate the data a bit. Consider Figure 4. Suppose that Cassie’s background beliefs include both that Lincoln was the 16th president and that Lincoln died in 1865, and suppose once

---

1It may in fact be so stated; one of the coauthors confesses to not (yet) having read it; the other has read it but does not recall whether it is so stated. It might suffice for van Inwagen’s example that it follow (logically) from what is stated in *War and Peace* that Napoleon is vain; no matter.
again that she reads in a narrative that Lincoln was re-elected in 1868. The processing “problem” here (it is not really a problem) is that we want the first of the reader’s two background beliefs to “migrate into” the story world. The reason that this is not a problem is that those first two background beliefs are the reader’s beliefs and the third is not. The first one (that Lincoln was 16th president) is both believed by the reader and is in the story world.

Consider Figure 1 again. If Cassie knows that she is reading a narrative, we want it to be the case that she believes (1) (that Washington was the first president), and we want both (1) and (2) (that he chopped down the cherry tree) to be in the story world. How do we accomplish this? Under the first solution, all propositions from the narrative will be placed in a story context. Under the second solution, we start with a story operator on (2). In general, we will put a story operator on all narrative predicates.

But then we face two problems: Background beliefs of the reader are normally brought to bear on understanding the story, as we saw in Figure 2. And we often come to learn (or, at least, come to have beliefs) about the real world from reading fictional narratives. Thus, we need to have two rules, which we will put roughly, but boldly, as follows:

(R1) Propositions outside the story space established by the story context or the story operator (i.e., antecedently believed by the reader) are assumed, when necessary, to hold within that story space by default but defeasibly.

(R2) Propositions inside the story space are assumed, when necessary, to hold outside that story space by default but defeasibly.

Some comments: The “when necessary” clause is there to prevent an explosion in the size of belief and story spaces. The migrations permitted by these two rules would only take place on an as-needed basis for understanding the story or for understanding the world around us. The “by default” clause is there for obvious reasons: We wouldn’t want to have Lincoln’s dying in 1865 migrate into a narrative in which he is re-elected in 1868. The “defeasibly” clause is there to undo any damage that might be done at a later point in the narrative if such a migration had taken place, innocently, at an earlier point. Rule (R1) (or such refinements of it as will, no doubt, be necessary as implementation of the theory proceeds) aids in our understanding of the story. Rule (R2) (or such refinements of it as will also, no doubt, be necessary as implementation of the theory proceeds) allows us to enlarge our views of the world from reading literature, while also allowing us to segregate our real-world-beliefs from our story-world beliefs. In this manner, we can facilitate the membrane whose semi-permeability allows us to understand narratives using our world knowledge, and to learn from narratives—indeed, to understand the real world in terms of narratives (cf. Bruner 1990).

We will close with three final remarks. First, to see how the story operator solves the problem with Lewis’s theory, look at Figure 5. (How it solves the problems with van Inwagen’s are left as exercises for the reader.) Second, in Figures 1–5, we have used the linguist’s triangle to hide irrelevant details; however, Figure 6 shows how the story operator looks in detail. Finally, two demo runs of preliminary implementations using SNeBR are presented in Rapaport 1991 and in a longer version of this paper (Rapaport & Shapiro 1996: 118–127).

3 CONCLUSION.

This brings to an end our essay in computational philosophy. We have explored knowledge-representation and reasoning issues surrounding fictional entities and their fictional (and non-fictional) properties, as well as their “interaction” with non-fictional entities. We have shown how a reader (human or machine) could read a narrative and construct and reason about her mental model of the story expressed by the narrative, and how information can selectively flow between general “real-world” knowledge and story-world knowledge.
4 ACKNOWLEDGMENTS.

The work presented here was done in collaboration with the members of the SNePS Research Group and the Center for Cognitive Science at State University of New York at Buffalo, to whom we are grateful for their contributions and comments, especially Jürgen Haas, Susan Haller, Johan Lammens, Sandra L. Peters, and Janyce M. Wiebe. This research was supported in part by the National Science Foundation under Grant IRI-8610517. Versions of this paper were presented by Rapaport at the 1989 Conference on Problems and Changes in the Concept of Predication (University of California Humanities Research Institute, University of California at Irvine) and the First Annual SNePS Workshop (SUNY Buffalo; Kumar 1990). Previous versions of this paper appeared in Rapaport 1990 and as part of Rapaport 1991. The present version is a slightly edited abridgment of Rapaport & Shapiro 1995.

5 REFERENCES.


