THE METAPHYSICS OF INTERPRETATION

By

ALLISON JILL HEPOLA

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Philosophy

written under the direction of

Peter Kivy and Dean Zimmerman

and approved by

_______________________________

_______________________________

_______________________________

_______________________________

_______________________________

____________________________________________________________________

____________________________________________________________________

____________________________________________________________________

____________________________________________________________________

____________________________________________________________________

____________________________________________________________________

New Brunswick, New Jersey

May 2011
ABSTRACT OF THE DISSERTATION

The Metaphysics of Interpretation

by ALLISON JILL HEPOLA

Dissertation Directors:
Peter Kivy and Dean Zimmerman

In *The Metaphysics of Interpretation*, I explore the ontological issues surrounding fictional characters, literary works, and literary interpretation. My central claim is that if one accepts a certain position on the ontology of fictional characters and literary works – artifactualism – then, under certain circumstances, the misinterpretation of a literary work will result in the full-fledged destruction of that work. Some related matters that are studied include: realism about fictional characters, artifactualism’s implications for the debate between textualism and constructivism, the seeming temporal “gappiness” of literary works, and the question of validity in interpretation.
ACKNOWLEDGEMENTS AND DEDICATION

Above all, I am indebted to my dissertation co-directors Peter Kivy and Dean Zimmerman for their unflagging support and guidance throughout this project and for their extensive comments on its text. Also I am grateful to my outside reader Amie Thomasson, first for her many helpful comments and suggestions on the manuscript, and secondly for her ground-breaking work in artifactualism that has made this all possible. I owe a special thanks to previous committee member and advisor Ted Sider for first suggesting the topic of fiction to me and for his guidance during the initial stages of this project.

It has been said that philosophy is fundamentally a dialogue, and it is difficult to imagine how this project could have taken form without the enormously helpful feedback I received in conversations with faculty and fellow graduate students. I would like to thank Barry Loewer in particular for his feedback on the project. I am especially grateful to faculty from the Rutgers English department – William Galperin, Jonathan Kramnick, and Henry Turner – for taking time out of their busy schedules to discuss philosophy and literature with me. This project was greatly enriched by their insights about interpretation and literary theory.

I also wish to thank the following people for their helpful comments: Saba Bazargan, Nick Beckstead, Tim Campbell, Ruth Chang, Janice Chik, Joe Corabi, Heather Demarest, Kate Devitt, Richard Dub, Jeff Glick, Gabe Greenberg, John Haldane, Thomas Hibbs, Doug Husak, Michael Johnson, Peter Klein, Mark Lee, Karen Lewis, Howard
McGary, Brian McLaughlin, Jeff McMahan, Jenny Nado, Anthony O’Hear, Bryan Parkhurst, Andrew Sepielli, Justin Sharber, Stephen Shortt, Holly Smith, Joshua Stuchlik, Meghan Sullivan, Nicholas Teh, Jason Turner, Candace Vogler, Evan Williams, participants in Barry Loewer’s dissertation seminar, participants in the Center for Cultural Analysis’ April 2010 conference on fiction, and participants in the Witherspoon Institute’s Aesthetics and Morality Seminar.

Additionally I must thank two professors from my undergraduate days at Notre Dame: Mike Loux and Neil Delaney. During the first semester of my freshman year, I took introduction to philosophy with Mike and an honors humanities seminar with Neil. In the former, I bristled against Mike’s allegation that non-existent objects like dragons had no properties (we were discussing the ontological argument); in the latter, Neil had us read Borges and encouraged us to ponder, among other things, the number of Don Quixote-s. Little did I know at the time, but seeds were being planted that very first semester of college that some ten years later would blossom into significant parts of my dissertation. It is delightful how philosophy sometimes works out that way.

Finally I must thank my family, especially my husband Andy Poker. Without his constant support and encouragement, his occasional proof-reading assistance, and his invaluable background in literature, none of this would have been possible. It is to him that this dissertation is dedicated.
# TABLE OF CONTENTS

Abstract of the Dissertation ........................................ ii
Acknowledgements and Dedication .................................. iii
Introduction ................................................................ 1

Chapter 1. Fictional Realism and Anti-Realism ................. 12
Chapter 2. Varieties of Fictional Realism ......................... 57
Chapter 3. Artifactualism and Interpretation ..................... 137
Chapter 4. Gappiness .................................................. 189
Chapter 5. Lingering Questions ...................................... 220

Bibliography ................................................................ 260
Curriculum Vitae .......................................................... 272
INTRODUCTION

In *The Metaphysics of Interpretation* I explore just that - the ontological issues surrounding the interpretation of literary works. I argue that if one embraces artifactualism, then in some situations the misinterpretation of a literary work results in the full-fledged destruction of that work.

Artifactualism is the claim that literary works – poems, novels, short stories, plays – are abstract artifacts. While literary works are artifacts, they are not concrete entities like books, hammers, or bicycles. And while literary works are abstracta, they are not eternal and necessary entities – like abstracta traditionally conceived – but rather contingent things that can be destroyed. Literary works are contingent abstracta, a rather broad category that includes many other familiar “cultural” entities from everyday life such as marriages, laws, institutions, and the like.

A literary work comes into existence when an author (or group of authors) composes a text with the intention of creating a literary work. Once created, a literary work can continue to exist even long after the death of its author. A literary work depends for its continued existence upon the existence of at least one physical copy of its text and at least one reader who can read the language the text is written in and - as I will argue - who can correctly interpret the text. If no such text or reader exists, then the literary work no longer exists.

Artifactualism was first advocated by Roman Ingarden, but was fully developed by Amie Thomasson. Thomasson is responsible for coining the term artifactualism,
among other things. I accept Thomasson’s version of artifactualism, and it is her theory of it that I analyze and extend in new directions over the course of this dissertation.

I came to accept Thomasson’s theory of literary works in a rather roundabout fashion. Thomasson is also an artifactualist regarding fictional characters; according to her, entities like Hamlet and Jane Eyre are abstract artifacts. At the outset of my dissertation, I was deeply interested in the ontology of fictional characters. As I researched the matter, Thomasson’s account of fictional characters was by far the most compelling. But Thomasson’s theory of fictional characters is only coherent within her overall theory of literary works. Fictional characters stand and fall, so to speak, with the literary works in which they are found. This is hardly surprising since there is a rather obvious connection between a fictional character and whatever novel, short story, or play it appears in. One cannot fully understand the ontology of fictional characters without also understanding the ontology of literary works. So it was quite natural for me to adopt an artifactualist stance towards literary works. Subsequently I discovered that the artifactualist account of literary works is quite interesting in its own right and has some startling implications for the issue of interpretation.

My project’s origin in the ontology of fictional characters is reflected in the first two chapters. Chapter 1 is devoted to a discussion of fictional realism. Fictional realism is the belief that fictional characters such as Hamlet and Sherlock Holmes are objects. This formulation may sound odd; realism is generally believed to involve claims about existence, not objects. However I have chosen to define fictional realism as a claim about objects rather than existence for a very important reason. Not all fictional realists believe that fictional entities exist. Notably, Meinongian theories of fiction claim that
fictional entities are non-existent. Given the important historical role Meinongianism has played in fictional realism, I think it is best to take Meinongian views into account when defining fictional realism as a whole. What all forms of fictional realism have in common is the assertion that fictional entities are *bona fide* objects with *bona fide* properties.

Before presenting the case for fictional realism in Chapter 1, I trace the historical development of the theory with a heavy emphasis on the works of Franz Brentano and Alexius Meinong. I believe Meinong should be considered the first fictional realist; he is the first person to specifically devote philosophical treatment to the ontology of fictional characters, concluding that they (as well as literary works) are non-existent objects. But it is fortunate that Meinong is not the last word in fictional realism, given the unappealing metaphysical baggage of his overall ontological theory. Having given Meinong his due credit, I turn towards a more contemporary case for fictional realism, which I term the “linguistic case” for fictional realism.

The linguistic case typically begins with the observation that many people engage in discourse about fictional characters. When we engage in discourse about fictional entities, we often say things like “Hamlet is a prince,” “Sherlock Holmes is a detective,” and “Hamlet was created by Shakespeare.” Intuitively it seems that such sentences have truth values; the sentence “Hamlet is a prince” is true while the sentence “Hamlet is a winged horse” is false.

Fictional realists argue that the best way to account for our intuitions about the truth values of these kinds of sentences is to simply accept that fictional entities are objects. The sentence “Hamlet is a prince” is true because there is an object, Hamlet, that
has a property – being a prince. Likewise the sentence “Hamlet is a winged horse” is false because there is an object, Hamlet, that lacks a property - being a winged horse. If a philosopher does not wish to postulate fictional entities as objects, perhaps for independent reasons such as metaphysical parsimony, she will end up committed to unappealing views such as the rejection of existential generalization or the counterintuitive position that sentences like “Hamlet is a prince” and “Hamlet is a winged horse” lack true values entirely.

After presenting the linguistic case for fictional realism, I consider some anti-realist responses and strategies. The most common anti-realist response to the linguistic case argues that all statements of predication involving fictional characters are implicitly prefixed with “according to the story” or some similar operator. Thus we are only committed to the existence of stories, not to the existence of fictional characters, to account for the truth values of such predications. I respond by admitting that most of our discourse involving fictional characters is implicitly prefixed with “according to the story” - no serious fictional realist would disagree - but that some sentences like “Hamlet was created by Shakespeare” and “Jane Eyre is a fictional character” cannot be prefixed in this manner. We still must postulate fictional characters as objects in order to account for the truth values of those sentences.

Finally in Chapter 1, I examine Kendall Walton’s pretense theory. Walton is an anti-realist about fictional characters, and his pretense theory attempts to provide a positive account of what is really going on when we make predicative statements about fictional characters. Walton’s theory raises some worthwhile points about the role of imagination and pretense in our experience of fictional literature. However pretense
theory has significant flaws, and it does not provide a satisfactory account of the truth values of certain sentences referring to fictional characters.

Once the case for fictional realism has been made, an important question naturally arises. If fictional characters are indeed objects, what kind of objects are they? Chapter 2 is devoted to answering this question. I survey several contemporary and historical versions of fictional realism with the aim of discovering which one most closely adheres to our ordinary and scholarly intuitions about fictional characters. I discuss Meinongian, neo-Meinongian, and Platonic theories of fictional realism, as well as idealist theories of fictional realism. I conclude that all of these forms of fictional realism have unacceptably counter-intuitive consequences. Meinongian fictional realism and Platonic fictional realism classify fictional characters as eternal and necessary entities. Authors do not create their characters, but rather discover them or “pick them out” from an infinite realm of eternal and necessary entities. I believe this sharply conflicts with our ordinary intuitions and scholarly practices surrounding fiction, wherein authors are viewed as akin to parents. Meanwhile idealist fictional realism fares no better. Because idealist fictional realism identifies fictional characters with authorial mental states or with individual readers’ mental states, the result is a world in which there are literally millions of Hamlets, millions of Jane Eyres, and the like.

Chapter 2 concludes with the introduction of artifactualist fictional realism. I discuss artifactualism’s origins in the work of Roman Ingarden before presenting Thomasson’s modern form of the theory. Pointing out how deftly artifactualism avoids the counter-intuitive consequences of Meinongian, Platonic, and Idealist forms of
fictional realism, I argue that one should adopt artifactualism if one wishes to be a
fictional realist.

In Chapter 3 I turn my attention towards the ontology of literary works and
present my central thesis. Thomasson argues that a fictional character is created by its
author but depends for its continued existence upon the existence of whatever literary
work it appears in. Hamlet and Ophelia were brought into existence by Shakespeare but
they still exist today because Hamlet exists. If Hamlet were to be destroyed then Hamlet
and Ophelia would be destroyed along with it. Since literary works play such an
important role in artifactualist fictional realism, it is natural to consider what kind of
objects literary works are and what conditions are required for the continued existence of
literary works.

According to Thomasson, literary works are also abstract artifacts. They come
into existence only through the creative actions of their authors. An author creates a
literary work when he composes a text with the intention of creating a literary work. I
note that under artifactualism literary works have their origins essentially. If by some
incredible coincidence, two authors were to create two completely identical texts, this
would result in two distinct literary works. Furthermore if a literary work had been
created by its author at a different time from when it was actually created, the resulting
work would also be a distinct literary work. For example, if James Joyce had written the
text of Ulysses in 1916 instead of 1922, this novel would have been different from the
Ulysses Joyce did, in fact, create in 1922.

Once created, a literary work depends for its continued existence upon the
existence of at least one physical copy of its text and at least one reader who can read the
language the text is written in and who has the “relevant background assumptions” required to understand the text. If any of these conditions are lacking, the work will become non-existent. If every physical copy of the work’s text is destroyed, then the work will no longer exist. If the language a work is written in becomes completely extinct (such as Etruscan has), then the work will no longer exist even if physical copies of the work’s text still exist. And if no person exists who has the proper background required to understand the text, then the work will also no longer exist, even if physical copies of the text still exist and even if those copies are written in a language easily read by many.

Thomasson does not explain what exactly is meant by a reader with the relevant background assumptions required to understand a text. I examine this intriguing condition closely in Chapter 3. I argue that a reader with the “relevant background assumptions” is best understood as a reader who can properly interpret the text. Thus for any literary work to enjoy continued existence, there must exist at least one reader who can provide the correct interpretation of the work. If a state of affairs arises where every single person misinterprets the work, and no person can properly interpret the work, then the work does not exist any longer. This will be the case even if multiple copies of the work’s text abound and many people have the language capabilities required to read the text.

I set aside the question of just how “wrong” a misinterpretation can be before it becomes destructive until Chapter 5. In Chapter 3 I rely on what I take to be very clear-cut and intuitive examples of gross misinterpretations - for instance, a reader who
believes Orwell’s *Animal Farm* is a straightforward story about barnyard animals rather than a satire of Stalinism.

For the remainder of Chapter 3 I examine in detail what artifactualism implies for the relationship between a literary work, its text, and its interpretations. Under artifactualism it seems that an interpretation of a work is, in some sense, prior to the literary work itself; literary works depend upon being properly interpreted to stay in existence. Now this is very similar to a view called constructivism that claims that whenever a reader interprets a text, he creates a literary work. According to constructivists, a literary work just is its interpretation (or a class of its interpretations). Contrasted with constructivism is a view known as textualism which claims that a literary work is identical to its text. *Pride and Prejudice* just is the sequence of words that begins “It is a truth universally acknowledged . . .”

Both textualism and constructivism enjoy intuitive support on different points. However both views are fraught with unacceptably counter-intuitive metaphysical consequences. Fortunately there is a solution to this dilemma – artifactualism. Artifactualism recognizes that both textualism and constructivism are correct about one thing – a work of literature’s existence is tightly bound up with the existence of its text and the existence of interpretations of it. But where textualists and constructivists see identity, artifactualism sees dependence. A literary work is not identical to its text or to its interpretation, but it depends upon both types of entities. Artifactualism functions as a via media between textualism and constructivism, accounting for the intuitions that favor both views while also avoiding the negative metaphysical consequences of both views.
Chapter 4 is devoted to a specific metaphysical problem that arises from the artifactualist theory of literary works. Literary works depend for their continued existence upon the existence of both physical objects (copies of texts) and human mental states. Human mental states, particularly those associated with understanding a language and being able to interpret literature properly, seem very metaphysically “unstable.” A language can be lost for centuries and then rediscovered at some future point; Ancient Egyptian was lost until the discovery of the Rosetta Stone. The kind of cultural knowledge required to properly interpret some literary works can also be lost, found, and lost again. If literary works depend on these unstable entities, then literary works seem prone to ontological “gappiness.” In other words, it seems a literary work can go in and out of existence. A literary work might come into existence in 300 B.C., go out of existence in 400 A.D. when the work’s language becomes extinct, and “pop back” into existence in 1800 A.D. when the work’s language is re-discovered.

Is this a problem? Gappiness is an undesirable feature in persons and spatio-temporal objects; it is difficult to understand how a person who exists from time $t_3$ until $t_4$ could be identical to a person who exists from $t_{18}$ until $t_{20}$. But literary works are abstracta, and abstracta might be more accommodating of gappiness than concrete objects and persons are. Thomasson herself does not believe it is problematic for a literary work to go in and out of existence over the course of time.

As I explain in Chapter 4, initially I was inclined to disagree with Thomasson. Imagine that a literary work is composed at $t_3$. At $t_4$ the work’s language is lost, and at $t_{18}$ the language is rediscovered. It would seem that the work then exists from $t_3$ until $t_4$ and then from $t_{18}$ until, say, $t_{20}$. Originally I believed that no artifactualist could say that


a literary work that exists from \( t3 \) until \( t4 \) is identical to a literary work that exists from \( t18 \) until \( t20 \). According to artifactualism, literary works have their origins essentially. A literary work that originates at \( t3 \) can never be identical to a literary work that originates at \( t18 \). To solve this problem, I pursued some a strategy that would eliminate the gaps by showing that if a language is lost between 400 and 1800 A.D. a literary work written in that language does not actually go out of existence during that time. Such a work only appears to go out of existence. This strategy ultimately fails, in fact it ultimately conflicts with other tenets of artifactualism, as I recount in detail.

Fortunately there is a solution. My initial concern about gappy literary works violating the “essentiality of origins” feature of artifactualism is unfounded. A literary work that exists from \( t3 \) until \( t4 \) and a literary work that exists from \( t18 \) until \( t20 \) would indeed be non-identical if the first work is created at \( t3 \) and the second work is created at \( t18 \). But is it really the case that a literary work is created at \( t18 \)? I conclude that no literary work is created at \( t18 \), because in order to create a literary work an author must compose a text with the intention of creating a literary work. But nothing of the sort happens at \( t18 \); rather a literary work “resumes” existence because its language is rediscovered and the text is interpreted properly. As I argued in Chapter 3, none of these activities create a literary work. These activities might create things like objects of perception, but an object of perception is not a literary work.

Chapter 5, the dissertation’s final chapter, addresses some lingering questions from previous chapters. I return to a matter raised in Chapter 3 of just how “wrong” a misinterpretation can be before it becomes potentially destructive of literary works. This is a very difficult question, and I do not claim to provide a definitive answer. Rather I
consider two different approaches that strike me as relatively promising: claiming that potentially destructive misinterpretations are characterized by failure to acknowledge a work’s essential properties and claiming that said misinterpretations are characterized by failure to identify a work’s genre.

Having presented my thesis, I conclude the dissertation by addressing a question that tends to nag at any philosopher who has completed a lengthy, detailed, and narrowly focused project – why should we care? I believe we should care very much about artifactualism and its account of literary interpretation. The destructive potential of misinterpretation is far from a hypothetical scenario simply used to illustrate a finely grained ontological point. Rather the destruction of literary works via misinterpretation can and does indeed happen. Using a real-life example, I demonstrate that ancient works and works that originate from oral tradition are particularly vulnerable to destruction due to misinterpretation. This is a startling conclusion, and in light of it we would do well to reassess many of our ordinary beliefs and scholarly practices surrounding ancient texts, folklore, and other literary works.
Fictional Realism and Anti-Realism

My project begins with a study of the ontology of fictional characters. Here in this first chapter, I will discuss fictional realism, the claim that fictional characters are *bona fide* objects. I will closely examine the reasons why both historic and contemporary philosophers have adopted fictional realism. I will also consider historic and contemporary critiques of fictional realism, with special attention paid to Kendall Walton’s pretense theory.

1.1 Fictional Works

Since this dissertation will deal extensively with fictional characters and fictional works, it is important to clarify what exactly is meant by these terms. A fictional work is a type of *narrative*. A narrative, broadly defined, is a description of events and things. Narratives include written descriptions of events and things (i.e., texts), spoken descriptions of events and things (such as anecdotes, lectures, or conversation), and visual depictions of events and things (movies, television programs, stage dramas, editorial cartoons, and the like).

Narratives can be either fictional or non-fictional. *Fictional works* are narratives that predominately concern events and things that are not real. *Non-fictional works* are
narratives that concern events and things that are real or are commonly believed to be real. These are not meant to be exhaustive definitions; I hope it can largely be taken for granted by all readers what kinds of works are fictional or non-fictional. Quintessential examples of non-fictional works are encyclopedia entries, journal articles, essays, biographies, memoirs, newspaper articles, anecdotes, scientific reports, philosophical treatises, and historical chronicles. Typical examples of fictional works are novels, short stories, poetry, stage dramas, operas, ballads, movies, and sitcoms.

Most philosophical discussion of fictional works has focused on fictional works that are in written form, such as novels and short stories; I too will confine my discussion to these types of fictional works. But it is important to note that many fictional works are not in written form or exclusively in written form. Dramas, operas, and movies all have visual and spoken components and some short stories, such as folktales, are never found in written form.

The distinction between a work of fiction and a work of non-fiction is not always clear as well. Non-fictional works sometimes are said to have “literary merit” and other aesthetic qualities associated with fictional works. The essays of Petrarch, some of Plato’s dialogues, and contemporary New Journalism pieces come to mind. Meanwhile some fictional works, particularly satires, parodies, and allegorical literature, appear to be largely concerned with real things and real events.

Mythology also presents an interesting blurring of the lines between fictional and non-fictional narratives. To modern readers, a work like the Aeneid is concerned with unreal things and unreal events. Vergil and his audience, on the other hand, viewed the Aeneid as a non-fictional narrative concerned with real things and real events in the
history of Rome not much different from Caesar’s narratives about the conquest of Gaul. Typically in philosophical treatment of fiction, mythological characters have been conflated with fictional characters; Pegasus and centaurs belong to the same category as Sherlock Holmes and Hamlet.¹

1.2 Fictional Entities

The concept of a fictional character is such a simple and basic concept that one really only needs to appeal to examples of fictional characters to illustrate it. Fictional characters are things like Sherlock Holmes, Scarlett O’Hara, Hamlet, the Cheshire Cat, and so on. Most often the term “fictional character” refers to fictional people like Hamlet and Jane Eyre or fictional person-like creatures like the Cheshire Cat who populate fictional works. But it is important to note that there are many other things that populate


fictional works that are not people or person-like. There are fictional inanimate objects such as the ring of Sauron, or the estate of Tara. One never hears the ring of Sauron referred to as a character in *The Lord of the Rings*, but it seems to function within that literary work in the same way as fictional characters like Frodo and Gandalf.

Understandably, almost all philosophical treatment of fiction has focused on fictional characters in the narrow sense - that is, on fictional people like Hamlet and Anna Karenina. Some fictional people, like Hamlet and Sherlock Holmes, are extremely well-known cultural icons. Philosophers of fiction can safely assume that the majority of their audience is familiar with Hamlet and his properties (being a prince, being Danish, having angst, etc) even if a large segment of their audience has never read *Hamlet*. Philosophers cannot always assume that their audience will be as familiar with things like Ophelia’s willow tree or Gertrude’s poisoned goblet.

Nonetheless there are other things within fictional works besides fictional people; these things function within works like fictional characters do. Ontological claims about fictional people like Hamlet will also apply to fictional non-people like the poisoned goblet. Because the term “fictional character” has such a narrow connotation, I will use the term *fictional entity* instead. Hamlet, Sherlock Holmes, and Anna Karenina as well as the ring of Sauron, Tara, and the poisoned goblet are all fictional entities. Fictional entity can be more rigorously defined as follows:

If a fictional work concerns an entity $x$, and if $x$ is not a real entity, then $x$ is a fictional entity.
The notion of a fictional work “concerning” an entity needs some elaboration. A
fictional work concerns an entity \( x \) when the fictional work is in some sense “about” \( x \).
This means that the fictional work may refer directly to \( x \), indicate a proper name for \( x \),
simply say something about \( x \), or attribute some property to \( x \).

In order to be a fictional entity, \( x \) must be the kind of entity that is only found
“within” a fictional work or works.\(^2\) \( X \) cannot be an occupant of the “real,” spatio-
temporal, non-fictional world. This is an important point, because many real world
entities are found within fictional works (or at least appear to be found within fictional
works) alongside purely fictional entities. Napoleon is found within *War and Peace*, the
Castel Sant’Angelo is found within *Tosca*, the city of London is found within the
Sherlock Holmes stories, and so on. Since London is an occupant of the real world, it is
not a fictional entity even though several works of fiction concern London.\(^3\)

1.3 Fictional Realism

In its most general formulation, *fictional realism* is the claim that fictional entities
are *objects*. This formulation may sound odd; realism is generally believed to involve
claims about existence, not objects. However I have chosen to define fictional realism as
a claim about objects rather than existence for a very important reason. Not all fictional

\(^2\) I use “within” in only the metaphorical sense. Strictly speaking, fictional entities cannot be “in” fictional
works and fictional works cannot have entities “within” them since neither fictional entities nor fictional
works are spatio-temporal objects.

\(^3\) There is controversy among philosophers of fiction about whether the London of the Sherlock Holmes
stories is identical to the real city of London. Parsons argues that the real London does not actually appear
in the Sherlock Holmes stories, but instead a “surrogate” of London – which is not identical to the real city
– appears in these stories (Parsons, *Nonexistent Objects*, 51-52, 182-185). This is an interesting issue, but
largely tangential to my project. Suffice it to say, the London in the Sherlock Holmes stories, whether it is
identical to the real city or a surrogate for it, is not a fictional entity in the same way that Holmes and
Watson are.
realists believe that fictional entities exist. Notably, Meinongian theories of fiction claim that fictional entities are non-existent. Given the important historical and contemporary role Meinongianism has played in fictional realism, it is important to take Meinongian views into account when defining fictional realism as a whole.

Meinongians and non-Meinongian fictional realists all agree that fictional entities are objects. I hope it can be taken for granted that “object” is a primitive philosophical concept that does not require too much elaboration on my part. When fictional realism asserts that fictional entities are objects, it is asserting that there is an ontology for fictional entities just as there is an ontology for any other kind of object. Fictional entities can be quantified over, just as any other kind of object can. Fictional entities can stand in relations to other fictional entities and to non-fictional entities. Above all, statements of predication referring to fictional entities have truth values.

Just what kind of objects fictional entities are is hotly debated among fictional realists. In Chapter 2 I will discuss different realist theories in greater detail. For now, it is important to note that there is one point where all fictional realists are in agreement, to the best of my knowledge. Fictional entities are abstract objects. They are not spatio-temporal objects, although they may stand in important relations to spatio-temporal objects.

Perhaps it is no surprise then that the history of fictional realism is deeply connected to more general concerns about the metaphysics of abstract objects. The next three sections will examine discuss motivations for fictional realism. The first motivation for fictional realism, which I have termed the “historical case,” is part of the larger debate
about objects, abstracta, existence, and non-existence that dominated metaphysics at the turn of the century.

1.4 The Historical Case for Fictional Realism

There does not appear to be any philosophical discussion of the ontology of fictional entities prior to the nineteenth century. It is puzzling why questions about fictional entities did not arise until relatively recently, given how philosophers have been interested in fiction – especially the genres of poetry and drama – since classical times.

Plato discusses poetry in several dialogues and famously condemns most of it in the Republic. Aristotle wrote an entire text, the Poetics, detailing his own theory of poetry. Nonetheless Plato and Aristotle both fail to discuss anything resembling fictional realism or the ontology of fictional entities. Plato is primarily concerned with the power poetry can wield over human emotions, the role of poetry in the moral education of youths, and the appropriate social position for poets in an ideal polis. Aristotle focuses his discussion in the Poetics on the different categories of poetry and how they may be distinguished; he is particularly interested in the unique characteristics of tragedy.

Both Plato and Aristotle note that an important part of a poet’s craft involves imitation and representation of real things and real emotions. This certainly has some implication for the ontology of fictional entities, but frustratingly neither Plato nor

---

5 Plato, Republic Book II 378b-383c, Book III 386-392b.
8 Aristotle, Poetics, Chapters 4, 6, 9, 25; Plato, Republic, Book III 392d, 394c-397b, Book X 598e-608c.
Aristotle elaborates on the matter in a way helpful to modern day metaphysicians of fiction. Plato and Aristotle’s remarks on poetry do not answer the kind of questions that interest modern fictional realists such as: is Lysistrata and unchanging, eternal form, or was she created in time by Aristophanes? Does Lysistrata have any existence whatsoever? Does she have an essence?

To the best of my knowledge, the first explicit treatment of the ontology of fictional entities is found in John Stuart Mill’s *A System of Logic*, published in 1843. Ontology was not Mill’s main concern in writing *A System of Logic*; his comments about the ontology of fictional entities are found in the context of a larger discussion about propositions. According to Mill, a proposition is particular type of sentence, “a portion of discourse in which a predicate is affirmed or denied of a subject.”

Every proposition has three necessary components: a subject, a predicate, and a copula. A copula is a word or set of words that signals that a sentence contains predication and thus is a proposition. “Is,” “is not,” and various forms of “to be” are the most common copulas. Mill notes that the copula “is” seems to do more than signal that certain sentences contain predication. It is commonly thought that sentences containing the copula “is” also imply the existence of their subjects.

It is apt to be supposed that the copula is much more than a mere sign of predication; that it also signifies existence. In the proposition, Socrates is just, it may seem to be implied not only that the quality just can be affirmed of Socrates, but moreover that Socrates is, that is to say, exists.

Mill admits that the word “is” is ambiguous leading many people to believe that “is” can be used as a copula and also as a predicate like “blue” or “just.” But Mill strenuously

---

10 Mill, *A System of Logic*, 104
rejects the view that “is” is a proper predicate. The popular belief that sentences containing “is” imply the existence of their subjects is quite mistaken. Hoping to disabuse his readers of this notion, Mill points out the logical consequences of this belief:

That the employment of it [is] as a copula does not necessarily include the affirmation of existence, appears from such a proposition as this, A centaur is a fiction of the poets; where it cannot possibly be implied that a centaur exists, since the proposition itself expressly asserts that the thing has no real existence.\footnote{Mill, \textit{A System of Logic}, 104.}

Mill appears to take it for granted that his readers believe that fictional entities such as centaurs do not exist. The existence of poetic creations like centaurs is so implausible that any theory of predication that leads to this conclusion can be dismissed on those grounds alone. Mill’s remarks about the proposition “a centaur is a fiction of the poets” seem to indicate a belief that “fictional” is equivalent to “non-existent.” The mere fact that centaurs are created by poets is sufficient to entail that centaurs do not exist. Centaurs are not fictional entities that happen to be non-existent; centaurs are non-existent in virtue of being fictional.

Elsewhere in \textit{A System of Logic}, Mill refers to other fictional entities and always notes that such things are non-existent. Most of the fictional entities Mill refers to – hobgoblins, dragons, hippocriffs, chimeras, mermaids, ghosts,\footnote{Mill, \textit{A System of Logic}, 66. 159, 200.} and Jupiter\footnote{Brentano, “Presentation and Judgment,” 67.} – are found in mythology and folklore but he does mention the Houyhnhms from \textit{Gulliver’s Travels} as another example of non-existent fictions.\footnote{Mill, \textit{A System of Logic}, 188.}
Mill’s claims about fictional entities do not appear to have attracted controversy at the time of *A System of Logic*’s publication. This suggests that by the mid-nineteenth century, the prevailing philosophical view regarding fictional entities was that such things were non-existent and deserved no further attention. The notion that centaurs and Houyhnhms could exist, let alone that philosophers could and should construct an ontology for them, would have been considered absurd.

But by the end of the nineteenth century, questions about the ontology of centaurs and other entities previously considered “non-existent” were at the forefront of metaphysics. This paradigm shift can be attributed to Franz Brentano (1838 - 1917).

Brentano did not set out to do metaphysics; his contributions to the ontology of fictional entities grew out of his general project of developing a philosophy of psychology. Psychology as an independent discipline began during Brentano’s lifetime, and it initially appeared to be confused, fragmented, and non-rigorous compared to the natural sciences. But Brentano believed that this new field of study had the exciting potential to become a full-fledged and respectable science, on a par with the natural sciences. Instead of taking physical phenomena as its area of study, psychology would be a complete “science of mental phenomena.”

Brentano realized that in order for psychology to become a full-fledged science, philosophers and psychologists would have to provide an explicit and rigorous account of the subject matter of psychology. Specifically, philosophers of psychology needed to provide a systematic account of the nature of mental phenomena and how mental phenomena are distinct from physical phenomena. Discovering and explaining the

---

unique character of mental phenomena was Brentano’s overriding purpose in writing his

The conventional way of distinguishing mental phenomena from physical
phenomena was to assert that mental phenomena were non-spatial, while physical
phenomena involved objects extended in time and space. Brentano did not reject this
distinction; but he believed it had become embroiled in too many controversies and
ambiguities to be adequate for his task.16 There were other features of human
consciousness besides non-extension that could better explain the fundamental difference
between mental and physical phenomena.

In *Psychology*, Brentano provided several different alternative criteria for
distinguishing mental from physical phenomena. The most famous criterion (as well as
the one that gave birth to modern fictional realism), is what became known as the “theory
of intentionality.”17 Brentano noted that whenever a person engages in psychological
activity – that is when a person thinks, believes, perceives, desires, judges, and the like –
the person’s mental activity is directed towards *some thing*. He writes, “in presentation
something is presented, in judgment something is affirmed or denied, in love [something
is] loved, in hate [something] hated, in desire [something] desired, etc.”18

As a result of this “directedness” of mental phenomena, every mental act contains

*something as object within itself.*19 The mental act of desire includes within itself the
object that is desired, the mental act of perception includes within itself the object that is

---

16 Franz Brentano, “The Distinction Between Mental and Physical Phenomena,” in *Realism and the
Background of Phenomenology*, ed. Roderick Chisholm (Atascadero, California: Ridgeview Publishing
17 It should be noted that Brentano himself never used this term.
19 Brentano, “The Distinction,” 50, emphasis added.
perceived, and so on. The importance of Brentano’s claim here cannot be overstated. If Brentano’s account of mental phenomena is correct, then there is an entirely new category of objects: the objects of mental activities. These objects are distinct from mental activities themselves; the cup of coffee I desire is not identical to my act of desiring a cup of coffee. Instead the cup of coffee that my mental activity is directed towards is an entirely separate object. Brentano believed that one must accept objects like this in order to make any sense of mental phenomena. Mental phenomena do not just happen to “contain” objects; the containing of objects is an essential property of mental phenomena. “We can define mental phenomena by saying that they are such phenomena as include an object intentionally within themselves.”

Brentano believed intentionality was the unambiguous criterion for distinguishing mental from physical phenomena that he was searching for. While perception, belief, and desire are always about something and contained that thing as an object, “no physical phenomenon manifests anything similar.” A tree is not about anything. A color does not include an object within itself.

However Brentano’s theory of intentionality provides yet another way of distinguishing mental from physical phenomena. Mental phenomena have a sort of freedom that physical phenomena do not; the objects of mental phenomena do not have to exist. It is immensely easy for anyone to have thoughts, beliefs, desires, judgments, or fears about non-existent objects. I can hold a belief about a centaur, fear a dragon, or desire to ride on Pegasus. As Chisholm notes in his discussion of Brentano, “we can

desire or think about horses that don’t exist, but we can ride on only those that do.”22 The fact that some mental activities can be directed towards non-existent objects sharply distinguishes mental from physical phenomena. “We may thus take it to be valid that the intentional inexistence of an object is a general distinguishing characteristic of mental phenomena, which differentiates this class of phenomena from the class of physical phenomena.”23

There are profound ontological consequences to Brentano’s recognition that many mental phenomena contained objects that did not exist. Since every mental event directed towards an object, there must be a distinct category of objects – the objects of mental activity. Some of the objects in this category will be non-existent, yet objects nonetheless.

In Brentano’s theory of intentionality, we see the beginnings of fictional realism. Assuming Mill’s views about fiction were dominant in the late nineteenth century, fictional entities such as centaurs and Pegasus would have been considered non-existent by Brentano and his contemporaries. However, a person can easily have beliefs, desires, and other mental activities directed towards Pegasus. Under Brentano’s theory of intentionality, all of those mental activities by definition contain an object – Pegasus. Pegasus may not exist, but that does not negate the fact that Pegasus is still an object of some sort. Pegasus must be an object in order for any person to have mental activity directed towards Pegasus.

Brentano does mention fictional entities as examples of non-existent objects contained in mental activities. He notes that mental phenomena can be directed towards

objects such as a blue swan, a learned bird, a living stone, an immortal man,\textsuperscript{24} and “a centaur [which] does not exist in a strict sense.”\textsuperscript{25} While a centaur does not exist, the theory of intentionality requires that a centaur still be an object of some sort in order for us to engage in mental activities directed towards centaurs. This is the case even when one’s mental activity regarding a centaur is to judge that a centaur does not exist. Brentano says that “if I make a judgment ‘A centaur does not exist,’ then it is said that \textit{the object is a centaur}, but that the content of the judgment is that a centaur does not exist, or the non-existence of a centaur.”\textsuperscript{26}

Interestingly Brentano appears to take a different position on the ontological status of fictional entities elsewhere in \textit{Psychology}. In the early 1870s, Brentano exchanged a series of letters with Mill. Sadly this correspondence ended with Mill’s sudden death in 1873, leaving many of Brentano’s questions and critiques unanswered.\textsuperscript{27} Brentano recounts one such letter he wrote critiquing Mill’s treatment of fictional entities from \textit{A System of Logic}

The proposition “A centaur is a poetic fiction,” requires (as you remark correctly) not that a centaur should exist, but the opposite. But in order for it to be true it requires that something else should exist, namely a poetic fiction which combines in a particular way the upper parts of a human body with the body and legs of a horse. If there were no poetic fiction and if there were no centaurs \textit{imaginatively created} by poets, then the proposition would be false; and its meaning is actually none other than that “there is a poetic fiction, which conceives the upper parts of a human body to be joined to the body and legs of a horse” or (which comes to the same thing) “There is a centaur \textit{imaginatively created} by poets.” The like holds true if I say that Jupiter is a Non-ens, i.e., that he is something which exists

\textsuperscript{24} Brentano, “Presentation and Judgment,” 64-66.
\textsuperscript{25} Brentano, “Genuine and Fictitious Objects,” 71.
\textsuperscript{26} Brentano, “Genuine and Fictitious Objects,” 71-72, emphasis added.
\textsuperscript{27} Brentano, “Presentation and Judgment,” 69.
merely in the imagination, not in reality. The truth of the proposition does not require that there should be a Jupiter, but that something else should exist.\footnote{Brentano, “Presentation and Judgment,” 68, emphasis in original.}

Here Brentano appears to contradict his position that fictional entities like centaurs are non-existent. Instead he seems to postulate a new category of objects, “poetic fictions.” These objects are distinguished by being “imaginatively created” by poets and, most importantly, these objects exist. If this is, in fact, what Brentano meant then his views on fiction in this passage anticipate a form of fictional realism known as artifactualism, which will receive considerable attention throughout this dissertation.

Since Brentano was ultimately not interested in providing a systematic ontology, his specific views on the metaphysics of fictional entities must remain the province of speculation. Nonetheless Brentano’s theory of intentionality laid the foundation for fictional realism. And while Brentano was ambiguous about the metaphysical implications of his new theory, his student Alexius Meinong (1853-1920) was more than willing to fill in the gaps and provide a systematic account of the ontology of non-existent objects.

The foundations of Meinong’s ontology are expressed in his 1904 work, \textit{The Theory of Objects}. At the beginning of this work, Meinong discusses Brentano’s theory of intentionality. Meinong strongly supported Brentano’s claim that mental activities are always directed towards some object.\footnote{Alexius Meinong, “The Theory of Objects,” in \textit{Realism and the Background of Phenomenology}, ed. Roderick Chisholm (Atascadero, Calif.: Ridgeview Publishing Company, 1960), 76.} He echoed Brentano in pointing out that the “directedness” of thoughts and beliefs provided a decisive criterion for distinguishing mental phenomena from physical phenomena.\footnote{Meinong, “The Theory of Objects,” 77.} And like Brentano, Meinong believed
the “something” a mental activity was directed towards is best classified as an object. He wrote that, “reference, indeed explicit directedness (Gerichtetsein), to that ‘something’ or (as we say quite naturally) to an object, unquestionably focuses itself upon our attention.” Unlike Brentano, Meinong was concerned with providing a systematic account of the ontology of objects of mental activities. His overriding purpose in *The Theory of Objects* was just that - to provide a rigorous metaphysical account of the objects of mental phenomena and of objects in general - in other words, a science of objects.

This new science of objects would have to be very broad and study both existent and non-existent objects. Meinong realized that any theory of objects that held non-existent objects, or unreal objects, to be worthy of serious study would clash with his contemporaries’ deeply held intuitions. He observed that humans are naturally inclined to take a “lively interest in reality” and the average philosopher, “tends to favor that exaggeration which finds the non-real a mere nothing, or . . . the non-real to be something for which science has no application at all or at least no application of any worth.” But Meinong was not deterred by such prejudices. Brentano was correct; every mental activity was directed towards an object. While many of our mental activities are directed towards real, existing objects, many other of our mental activities are directed towards objects that do not exist. One can hold beliefs about centaurs, fear dragons, or desire to ride on Pegasus. This basic fact about human psychology convinced Meinong that “there is not the slightest doubt that what is supposed to be the Object of

knowledge need not exist at all.”

Given the astounding power of the human imagination, Meinong believed that the amount of non-existent objects of mental phenomena (or “objects of knowledge” as he called them) would be quite vast indeed. “The totality of what exists, including what has existed and will exist, is infinitely small in comparison with the totality of the Objects of knowledge.”

Meinong realized that he could not simply postulate endless numbers of non-existent objects. He needed to provide a thorough account of just what it meant for an object to be “non-existent.” Such an account would have to begin with an examination of being or Sein.

According to Meinong, there are two distinct forms of being – existence and subsistence. Existence is the form of being that spatio-temporal objects have. Cats exist, trees exist, Rome exists, and so on. Meanwhile, subsistence is the form of being that certain abstract objects have. Mathematical and geometric objects like numbers, a circle, a straight line, and an angle as well as logical concepts like similarity and difference all subsist.

A very common misconception is that Meinong believed non-existent things like centaurs subsist. This misconception seems to arise from two sources: Meinong treats the terms existence and subsistence as antonyms, and he describes abstract entities as subsisting. But in reality, Meinong denied that non-existent “objects of knowledge” like centaurs and Pegasus subsisted.

Our account up to now may seem to leave room for the conjecture that whenever existence is absent, it not only can be but must be replaced by subsistence. But even this restriction is inadmissible.\(^{38}\)

According to Meinong, there is a large category of objects that neither exist nor subsist. Such objects lack being entirely. Some of the objects that are “outside of being” are objects that could possibly exist, such as the famous golden mountain. Some of the objects that are outside of being are objects that are impossible by their very nature, such as a round square.\(^{39}\)

What cannot be stressed enough is that under Meinong’s theory, objects that lack being still are objects. Not only are there objects that lack being, these objects also have genuine properties. Meinong referred to an object’s having properties as Sosein. The golden mountain has the properties of being golden and being a mountain. The round square has the properties of being round and being square. The proposition, “the golden mountain is made of chocolate” is false because it attributes a property to an object that the object does not, in fact, have. The fact that the golden mountain has no being whatsoever is irrelevant to the fact that it is still an object that genuinely has and lacks various properties. As Meinong put it, “the Sosein of an Object is not affected by its Nichtsein [its non-being]. The fact is sufficiently important to be explicitly formulated as the principle of the independence of Sosein from Sein.”\(^{40}\)

And while the claim that there are objects that have no being whatsoever undoubtedly strikes many people as bizarre and contradictory, Meinong argued that we must postulate such objects due to the theory of intentionality. We would not be able to

---

\(^{39}\) Meinong, “The Theory of Objects,” 82.
\(^{40}\) Meinong, “The Theory of Objects,” 82.
make the obviously true judgment that a round square is an impossible object unless there is an object, the round square.

Any particular thing that isn’t real (Nichtseidendes) must at least be capable of serving as the Object for those judgments which grasp its Nichtsein. It does not matter whether this Nichtsein is necessary or merely factual; nor does it matter in the first case whether the necessity stems from the essence of the Object or whether it stems from aspects which are external to the object in question. In order to know that there is no round square, I must make a judgment about the round square.41

The Theory of Objects was intended as a very general explication of Meinong’s ontology, and he did not provide other examples beyond the golden mountain and the round square of objects that lack being entirely. And notably for our purposes, Meinong says nothing about fictional entities and what kind of objects they might be. However, shortly after the publication of The Theory of Objects, Meinong had the opportunity to apply his ontology to fictional entities.

In 1905 the philosopher and psychologist Theodor Lipps published an essay titled, “Weiteres zur ‘Eingehlung,’” or “Further Considerations on Empathy.”42 In the essay Lipps discusses various “aesthetic objects,” including fictional entities like Goethe’s Mephistus. Lipps was concerned with how readers of Goethe’s text can make judgments about Mephistus and have “aesthetic feelings and pleasures” towards Mephistus.43 This phenomenon struck Lipps as quite interesting because the people who read about Mephistus are standard, spatio-temporal, existing, objects while Mephistus himself is a very unusual sort of object.

42 Raspa, “Fictional and Aesthetic Objects,” 52.
43 Raspa, “Fictional and Aesthetic Objects,” 52.
The Mephistus of the story . . . has a peculiar form of existence. Without any doubt he was once called into existence by Goethe. But having been called into existence and having reached artistic representation in the words of literature, he has a type of reality; what he does and says is in a certain sense beyond any doubt a verifiable “fact.”

The “type of reality” that Lipps believed fictional entities like Mephistus had was a unique kind of reality called “aesthetic reality.” Aesthetic reality is distinct from the kind of reality spatio-temporal objects have. Aesthetic reality, according to Lipps, “corresponds to being thought.”

Meinong soon responded to Lipps with the essay “Über Urteilsgefühle: was sie sind und was sie nicht sind” or “Judgment-feelings: what they are and what they are not.” Here, Meinong wholeheartedly rejected Lipps’ notion of aesthetic reality. He wrote, “there is only one reality, that of the empirical world, and an ‘aesthetic reality’ that is outside of it is not reality at all.” Existence or reality, as Meinong defined it, was not an attribute that fictional entities like Mephistus could ever have. Only spatio-temporal objects could be said to exist or to be real, and fictional entities, according to Meinong, lacked temporal properties entirely.

The action of a drama, even of a historical drama, does not have any strict time placement, at least not in the absolute time, while it (the action) no more lacks relative time determination than its characters lack the past and also the future.

Since fictional entities like Mephistus do not and cannot exist, are they objects that subsist or objects that lack being entirely? Meinong provides an answer to this while

---

44 Lipps, quoted in Raspa, “Fictional and Aesthetic Objects,” 52.
45 Raspa, “Fictional and Aesthetic Objects,” 53.
46 Raspa, “Fictional and Aesthetic Objects,” 52.
47 Meinong, quoted in Raspa, “Fictional and Aesthetic Objects,” 53.
48 Meinong, quoted in Raspa, “Fictional and Aesthetic Objects,” 55.
discussing the creative activity of an artist later in “Judgment-feelings.” He observes that when an artist such as Goethe writes of Mephistus for the first time, “the aesthetic object [Mephistus] is picked out from among the infinite totality of the objects outside being.” In other words, fictional entities are objects which lack being entirely. They neither exist nor subsist. Nonetheless they are still objects and still have properties.

### 1.5 Problems with the Historical Case for Fictional Realism

While Meinong’s theory of non-existent objects was the genesis of fictional realism, it also was the genesis of fictional anti-realism. Fictional anti-realism is the position that fictional entities are not objects, do not require an ontology, and should not be admitted to any ontology already in use. The oldest motivation for fictional anti-realism is distaste for Meinong’s theory of objects, especially as articulated by Bertrand Russell.\(^{50}\)

Meinong had argued that fictional entities were objects, but objects of a very unusual sort; they were objects with properties but without being. Fictional entities were not the only kinds of objects that lacked being; Meinong’s theory also postulated objects such as the golden mountain and impossible objects such as the round square. Meinong believed that this unusual collection of objects deserved a place within our ontology. Not surprisingly, this claim attracted great controversy and the attention of Russell. Russell

---

\(^{49}\) Meinong, quoted in Raspa, “Fictional and Aesthetic Objects,” 56, emphasis added.

argued that Meinong’s non-existent objects were incoherent and contradictory. In “On Denoting,” Russell correctly notes that Meinong’s theory of objects:

. . . regards any grammatically correct denoting phrase as standing for an object. Thus ‘the present King of France,’ ‘the round square,’ etc., are supposed to be genuine objects. It is admitted that such objects do not subsist, but nevertheless they are supposed to be objects. 51

Russell feels that the notion of a being-less object is “in itself, a difficult view.” 52 But the fatal defect of Meinong’s theory is that it postulates objects that “are apt to infringe upon the law of contradiction.” 53 Meinong’s ontology allows patently impossible objects such as the round square which “is round, and also not round,” an utterly unacceptable contradiction as far as Russell is concerned. 54 Even Meinong’s non-existent objects that did not seem prima facie impossible, such as the golden mountain or the present King of France, ultimately collapsed into a puddle of hopeless contradiction.

For example, according to Meinongianism, there is an object denoted by the phrase “the present King of France.” This object presumably has the properties of being a king, being the King of France, being the King of France in 1905, etc. This object also lacks being entirely (that is, it neither exists nor subsists), but that has no bearing on the fact that the present King of France still has properties, due to Meinong’s principle of the independence of Sein and Sosein. The metaphysics underlying all of this may be dubious, but so far the present King of France does not appear to violate the law of non-contradiction.

However, just as there is an object denoted by the phrase “the present King of France,” Russell points out that under Meinong’s theory there is also an object denoted by the phrase “the existent King of France.”55 Now both Russell and Meinong agree that the existent King of France clearly does not exist. Objects must be spatio-temporal in order to exist, according to Meinongianism, and the existent King of France is not spatio-temporal. Nor is the existent King of France the kind of object that subsists; Meinong limits subsistence only to mathematical and logical objects. Meinong would place the existent King of France in the same category as the present King of France, the round square, the golden mountain, and Mephistus - objects that lack being entirely. But if this is the case then the existent King of France violates the law of non-contradiction. The existent King of France has the property of being existent but also has the property of being non-existent (because he lacks being entirely and existence is a type of being). The existent King of France is as contradictory as the round square.

According to Russell, the notion of an object that violated the law of non-contradiction is “intolerable.” A theory, such as Meinong’s theory of objects, that requires the postulation of contradictory objects ought to be rejected on those grounds alone. “If any theory can be found to avoid this result, it is surely to be preferred.”56 Russell, of course, believed his own theory of denoting avoided the incoherence of Meinong’s theory of objects. The theory of denoting dispensed with Meinong’s vast category of being-less objects. “With our theory of denoting,” Russell proclaims, “we are able to hold that there are no unreal individuals; so that the null class is the class

containing no members, not the class containing as members all unreal individuals.”

Fictional anti-realism was one consequence of Russell’s elimination of being-less objects. Fictional entities like Hamlet and Apollo were now banished along with the round square, the even prime other than two, the present King of France, and the golden mountain to the “realm of non-entities.”

We see traces of Russell’s critique of Meinong in contemporary arguments for fictional anti-realism. Russell’s criticisms of Meinong can be generalized as such: if Meinong was correct about the nature of fictional entities, fictional entities are objects. These objects can easily be shown to violate the law of non-contradiction. Consequently if Meinong is correct, there are objects which violate the law of non-contradiction, which is an intolerable conclusion that ought to be avoided at all costs. Likewise many contemporary fictional anti-realists point out that if fictional realism in general is true, and fictional entities are objects, then there are objects that violate the law of non-contradiction, the law of excluded middle, and other logical laws. This is not just a problem for Meinongian theories of fiction; it is a problem faced by any account of fiction that asserts fictional entities to be objects.

Anti-realists point out an obvious fact about how people generally think and speak about fictional entities. According to general intuitions about fictional entities, fictional entities have various properties. Hamlet has the property of being a prince, the property of being Danish, the property of stabbing Polonius, and so on. Sherlock Holmes has the property of being male, the property of being a detective, the property of living at 221B Baker Street in London, etc. Lady MacBeth has the property of being female, the

---

property of being malicious, the property of hallucinating a dagger, etc. Why does nearly everyone believe that these fictional entities have these specific properties? For a simple reason: the fictional works that contain Hamlet, Sherlock Holmes, and Lady MacBeth ascribe these properties to these fictional entities. Sherlock Holmes has the property of living at 221B Baker Street because the series of Sherlock Holmes stories explicitly state that he has this property. Lady MacBeth has the property of being malicious because it can be reasonably inferred from the way she is described by MacBeth that she has this property.

It would seem then that if we take commonsense intuitions about fiction seriously, fictional entities have properties, and they have all the properties that are ascribed to them by their respective stories. It is important to note that fictional entities apparently have properties that are not ascribed to them by their stories; Hamlet has the property of being created by Shakespeare even though the text of Hamlet never ascribes this property to Hamlet. But for every property that Hamlet ascribes to Hamlet, Hamlet the fictional entity truly and genuinely has this property.

While this view of fictional entities and their properties appears to cohere with commonsense intuitions about fiction, anti-realists argue that it leads to devastating consequences for fictional realism. Fictional entities are objects, according to fictional realism, and these objects have various properties which are determined by how these objects are described by their respective stories. However it is not difficult to find examples of fictional works that ascribe contradictory properties to fictional entities or describe fictional entities in such a way that these entities become impossible objects.
For example, the Sherlock Holmes stories contain a character named Watson. At one point in the Sherlock Holmes series, Watson is given the property of having his (presumably singular) war wound on his leg. At another point in the series, the very same character Watson is ascribed the property of having his war wound on his arm.\(^{59}\) If fictional realism is true then Watson is an object. Watson has the properties ascribed to him by his stories. Therefore Watson is an object with the property of having a singular wound on his arm and the property of having a singular wound on his leg. This is a violation of the law of non-contradiction. Fictional realism has led to a very unappealing result - the postulation of an object that violates the law of non-contradiction.

Other fictional entities violate the law of excluded middle in virtue of how they are described in their respective stories. A famous case is Lady MacBeth and the number of children she had.\(^{60}\) According to *MacBeth*, Lady MacBeth has the property of having given birth.\(^{61}\) But the text of *MacBeth* never indicates the specific number of children Lady MacBeth gave birth to. Nor are there any reasonable interpretations of *MacBeth* that would indicate the specific number of children Lady MacBeth gave birth to. Lady MacBeth may have the property of having given birth to less than two children, or she may have the property of having given birth to two or more children. There is no way to tell which of these properties she has given what the text says about her. Yet the only way Lady MacBeth could have the property of having given birth to less than two

\(^{59}\) Parsons, *Nonexistent Objects*, 184.


\(^{61}\) “I have given suck and know how tender ‘tis to love the babe that milks me.” *MacBeth*, Act 1 Scene 7.
children is if the text ascribes this property to her. It seems that Lady MacBeth has been left hopelessly indeterminate with regards to the property of having given birth to less than two children. But if this is so, fictional anti-realists point out, then Lady MacBeth violates the law of excluded middle. An object cannot lack the property of having given birth to less than two children and lack the property of having given birth to two or more children. Since fictional realism requires Lady MacBeth to be an object, fictional realism has once again postulated an object that violates a basic law of logic.

Fictional entities can also display vagueness of identity. In “Against Fictional Realism,” Anthony Everett points out how easy it is to write a story containing fictional entities that violate these logical laws. He presents a story of his own, entitled Frackworld, that contains two fictional entities, Frick and Frack. Frackworld clearly states that it is vague whether Frick has the property of being identical to Frack, whether Frack has the property of being identical to Frick, or whether Frick and Frack have the property of being non-identical to one another.62 This presents a serious problem for fictional realism, because fictional realism is committed to Frick and Frack being objects. Since Frick and Frack have all the properties ascribed to them by their story, and their story ascribes to them the property of indeterminate identity, Frick and Frack are objects that display indeterminacy of identity. Appealing to Gareth Evans’ famous argument,63 Everett points out that it is not possible for any object to display indeterminacy of identity.64

A second story penned by Everett, Asymmetryville, contains the fictional entities of Cicero and Tully. The story explicitly ascribes Cicero the property of being identical

64 Everett, “Against Fictional Realism,” 629-630.
to Tully, while Tully is explicitly ascribed the property of being distinct from Cicero.\textsuperscript{65}

The fictional entities of Cicero and Tully violate the law of symmetry of identity. Once again this poses a serious problem for fictional realism. Fictional realism must postulate Cicero and Tully as objects that have all the properties ascribed to them by their story *Asymmetryville*. Since *Asymmetryville* says that Cicero and Tully violate the law of symmetry of identity, fictional realists must accept that there are objects that violate the law of symmetry of identity.

Fictional anti-realists believe that these kinds of cases are more than sufficient to show that fictional entities are not objects, or are at least not the kind of objects any respectable metaphysician would want to populate her ontology. Fictional entities are far too vulnerable to contradictions and violations of logical laws to be considered *bona fide* objects.

1.6 The Linguistic Case for Fictional Realism

The historical motivation for fictional realism, born from Brentano’s theory of intentionality and refined by Meinong’s ontology, is still employed by some contemporary fictional realists.\textsuperscript{66} But it is by no means the dominant form of fictional

\textsuperscript{65} Everett, “Against Fictional Realism,” 634.

realism nowadays. It is not difficult to understand why many contemporary fictional
realists would shy away from the historical case for fictional realism. The historical case,
due to its origins in Brentano, commits its adherents to a specific set of views about
psychology and the nature of mental phenomena; such views are hardly uncontroversial.
Historical fictional realism is also closely bound to Meinong’s theory of objects. Given
the devastating critiques leveled against Meinong’s theory, it is quite understandable that
many contemporary metaphysicians wish to avoid positions that entail Meinongianism.

In light of the difficulties faced by the historical case for fictional realism, a new
defense of fictional realism arose. I will refer to this motivation for fictional realism as
the “linguistic case” for fictional realism. Compared to the historical case for fictional
realism, the linguistic case is elegantly simple. It relies on commonly held intuitions and
avoids the messier ontological commitments of the historical case. Not surprisingly, the
linguistic case has become the predominate defense of fictional realism in contemporary
metaphysics.67

The linguistic case for realism takes as its starting point the simple observation
that people engage in discourse about fictional works and fictional entities. This

---

67 See David Braun, “Empty Names, Fictional Names, Mythical Names,” Nous 39, no. 4 (2005), 609-610;
Crittenden, Unreality, 95; Everett, “Against Fictional Realism,” 624-626; Everett, “Referentialism and
Empty Names,” in Empty Names, Fiction, and the Puzzles of Non-Existence, ed. Anthony Everett and
Thomas Hofweber (Stanford, Calif.: CSLI Publications, 2000), 38-39; Stacie Friend, “Real People in
Unreal Contexts,” in Everett and Hofweber, 184-186; Howell, “Fictional Objects: How They Are and How
They Aren’t,” 152; Frederick Kroon, “Disavowal Through Commitment: Theories of Negative
Existentials,” in Everett and Hofweber, 95; Parsons, Nonexistent Objects, 52-54; Plantinga, The Nature of
Necessity, 153; Nathan Salmon, “Nonexistence,” Nous 32, no. 3 (1998), 277; Thomasson, Fiction and
Metaphysics, 5-6; Thomasson, “Speaking of Fictional Characters,” Dialectica 57, no. 2 (2003); Quine, “On
What There Is,” 2; Kenneth Taylor, “Emptiness Without Compromise,” in Everett and Hofweber, 17-18;
Peter van Inwagen, “Creatures of Fiction,” in Ontology, Identity, and Modality: Essays in Metaphysics
(Cambridge: Cambridge University Press, 2001), 43; van Inwagen “Existence, Ontological Commitment,
and Fictional Entities,” 136-138; van Inwagen, “Quantification and Fictional Discourse,” in Everett and
Hofweber, 243-244; Zalta, “The Road Between Pretense Theory and Abstract Object Theory,” 118-120.
discourse can range from ordinary people discussing the latest hit movie or best-selling beach novel to the advanced literary theory found in English departments. Whether discourse about fiction takes place at the level of the folk or of the experts, it is largely made up of sentences that make reference to fictional entities and their properties, such as:

- Hamlet is a fictional character.
- Hamlet is a prince.
- Lady MacBeth is malicious.
- Lady MacBeth is more malicious than Miranda.

Discourse about fiction also contains sentences that describe relationships between fictional entities and real life entities, such as:

- Hamlet was created by William Shakespeare.
- Frodo Baggins is shorter than Michael Jordan.
- I am thinking about Lady MacBeth.

The majority of people have the intuition that these sentences and others like them are true. If one is inclined to be skeptical about the truth of sentences like “Hamlet is a prince,” it is helpful to consider sentences of the following sort:

- Hamlet is a winged horse.
- Lady MacBeth was created by Jane Austen.
- Miranda is more malicious than Lady MacBeth.
- Frodo Baggins is taller than Michael Jordan.

Such sentences strike nearly everyone as patently false. The first two sets of sentences seem far closer to “the truth” than the third set of sentences does. Fictional realists
believe that the best way to account for such intuitions is to simply say that sentences of the first sort are true and sentences of the second sort are false.

Fictional realists argue that the fact that sentences like “Frodo Baggins is very short” and “Hamlet was created by Shakespeare” have clear-cut truth values commits us to the existence of fictional entities. If there are no such things as Lady MacBeth and Miranda, it is difficult to see how it could still be true that “Lady MacBeth is more malicious than Miranda.” If there was no such thing as Frodo Baggins, then it is difficult to account for why the sentence “Frodo Baggins is shorter than Michael Jordan” seems so much “more true” than the sentence “Frodo Baggins is taller than Michael Jordan.” Why does the latter sentence sound so “wrong” to begin with if there is no such thing as Frodo Baggins to have or lack the property of being taller than Michael Jordan?

There are a few different ways a fictional realist may support the linguistic case for realism. Maria Reicher notes how simple statements of predication involving fictional entities (e.g., “Pegasus is a winged horse”) are enough to entail the existence of fictional entities. The sentence “Pegasus is a winged horse” is intuitively true. When combined with the principle of existential generalization (Fa → there is at least one F), this sentence leads to the conclusion that “there is at least one winged horse.”68 The sentence “Pegasus is a winged horse” may also be combined with the predication principle (Fa → a exists) to reach the conclusion “Pegasus exists.”69

Reicher argues that as long as one accepts that “Pegasus is a winged horse” is true, he will be committed to the existence of a fictional entity, Pegasus. If one does not wish to admit fictional entities into his ontology, he is left with two unappealing choices.

69 Reicher, “Two Interpretations,” 154.
He can take the radical step of rejecting existential generalization and the predication principle, or he can adopt the counterintuitive position that sentences like “Pegasus is a winged horse” and “Hamlet was created by Shakespeare” are not true.⁷⁰

Other fictional realists appeal to sentences about fictional entities that are more complex than simple statements of predication. These realists note that sometimes valid inferences can be constructed out of structurally complex sentences that refer to fictional entities. This strategy has notably been employed by Peter van Inwagen.⁷¹ Van Inwagen presents us with a sentence that clearly refers to fictional entities:

There is a fictional character who, for every novel, either appears in that novel or is a model for a character who does.⁷²

Van Inwagen does not believe that this sentence is true, but that is beside the point. What is relevant is that from the first sentence we can validly deduce this next sentence:

If no character appears in every novel, then some character is modeled on another character.⁷³

The fact that a valid inference can be constructed from these two sentences can only be accounted for if we postulate fictional characters. Van Inwagen observes:

The second sentence is a formal consequence of the first . . . now note a second formal consequence of the first sentence: ꞈx, x is a fictional character - that is to say: “There are fictional characters.” It seems, therefore, that the logical relations

⁷⁰ Reicher, “Two Interpretations,” 154.
⁷¹ See van Inwagen, “Existence, Ontological Commitment, and Fictional Entities,” and “Quantification and Fictional Discourse.”
⁷² van Inwagen, “Quantification and Fictional Discourse,” 243.
⁷³ van Inwagen, “Quantification and Fictional Discourse,” 244.
among certain sentences of fictional discourse can be accounted for only on the assumption that there are fictional characters.\textsuperscript{74}

The linguistic case for fictional realism is relatively simple and quite compelling. It has the added advantage of avoiding commitments to Brentano’s psychology or Meinong’s ontology. And it is very difficult for fictional anti-realists to refute. Even if the anti-realists are correct and Lady MacBeth is a logically incoherent object, it still seems that there are true sentences about her such as, “Lady MacBeth is malicious” and “Lady MacBeth is female.” It also seems that there are sentences about Lady MacBeth that are blatantly false such as, “Lady MacBeth is a winged horse” and “Lady MacBeth stabbed Polonius.” It is difficult to understand why sentences like “Lady MacBeth is malicious” are true and why sentences like “Lady MacBeth is a winged horse” are false without postulating Lady MacBeth as some kind of object.

Some fictional anti-realists, notably Russell and Frege,\textsuperscript{75} are willing to cast aside commonsense views about fiction and assert that sentences like, “Lady MacBeth is female” are simply false or lack truth values entirely. But given how much Russell and Frege’s approaches clash with widely held intuitions and beliefs about fiction, most fictional anti-realists have been reluctant to take them up. Instead fictional anti-realism has developed other strategies that preserve commonsense intuitions about the truth values of certain sentences about fiction but at the same time do not lead to the postulation of fictional entities as objects.

One common anti-realist strategy is to assert that all sentences that make reference to fictional entities need to be seen as implicitly prefixed with “according to the

\textsuperscript{74} van Inwagen, “Existence, Ontological Commitment and Fictional Entities,” 137.
story” or some similar operator. In other words, the sentence “Lady MacBeth is malicious” ought to be read as “according to the story, [MacBeth] Lady MacBeth is malicious.” The sentence “Sherlock Holmes lives at 221B Baker Street” ought to be read as, “according to the story [the various Sherlock Holmes stories], Sherlock Holmes lives at 221B Baker Street.”

By making use of the “according to the story” operator, fictional anti-realists are able to preserve our intuitions that such sentences about fiction have truth values. We believe that “Lady MacBeth is malicious” expresses a truth. Prefixing this sentence with the “according to the story” operator does not conflict with our intuitions in the slightest. It is certainly true that according to the story, MacBeth, Lady MacBeth is malicious. In fact, as I mentioned earlier, the only reason one would attribute the property of malice to Lady MacBeth is because of the way she is described in MacBeth.

Likewise, the use of the “according to the story” operator can explain why sentences like “Lady MacBeth is a winged horse” strike most people as patently false. Since the sentence “Lady MacBeth is a winged horse” makes reference to a fictional entity, it must be read as “according to the story [MacBeth], Lady MacBeth is a winged horse.” The story MacBeth, of course, says nothing that even remotely suggests that Lady MacBeth is a winged horse; rather MacBeth states the opposite - Lady MacBeth is a human being. Since Lady MacBeth is not a winged horse according to the story, the sentence “Lady MacBeth is a winged horse” is false.

While sentences prefixed with “according to the story” preserve commonsense intuitions about the truth values of statements about fiction, these sentences do not require an ontological commitment to fictional entities as objects. “According to the story, Lady MacBeth is malicious” is a true sentence, and its truth does not depend on there being a *bona fide* object, Lady MacBeth, that possesses the property of malice. All that is needed for “according to the story, Lady MacBeth is malicious” to be true is for there to be some story that at some point asserts (explicitly or implicitly) the proposition “Lady MacBeth is malicious.” This means we must postulate *fictional works* as objects in order to account for true sentences about fiction. But that is a far cry from fictional realism. It is relatively uncontroversial that fictional works like *MacBeth* and *Hamlet* exist and are have a place in our ontology.\textsuperscript{77} Controversy only arises when fictional realists argue that in addition to *MacBeth* and *Hamlet*, there are also objects like Lady MacBeth, Hamlet, and Polonius.

The use of “according to the story” as an implicit prefix certainly succeeds in allowing fictional anti-realists to avoid postulating fictional entities as objects while respecting commonsense intuitions about the truth values of some sentences about fiction. Unfortunately for fictional anti-realism, the use of this prefix is effective for only a limited number of sentences about fiction. There are many other sentences about fiction that intuitively seem true or false but which cannot be prefixed with “according to the story.”\textsuperscript{78} For example:

\textsuperscript{77} Thomasson, *Fiction and Metaphysics*, 139.
\textsuperscript{78} This weakness of the “according to the story” prefix is discussed in Howell, “Fictional Objects: How They Are and How They Aren’t” 152-153; Lewis, “Truth in Fiction” 263; Reicher, “Two Interpretations,” 154-159; Ryle, “Systematically Misleading Expressions,” 149; Thomasson, *Fiction and Metaphysics*, 95-97; Zalta, “The Road Between Pretense Theory and Abstract Object Theory,” 130.
Hamlet was created by Shakespeare
Hamlet was created by Jane Austen

The first sentence seems quite true. However, if it is prefixed with “according to the story,” the truth value of this sentence changes entirely. “According to the story, Hamlet, Hamlet was created by Shakespeare” is false. According to *Hamlet*, Hamlet was not created by Shakespeare. Instead Hamlet was created by his parents, Gertrude and the murdered king. No entity that remotely resembles William Shakespeare is found anywhere in the text of *Hamlet*.

The second sentence, “Hamlet was created by Jane Austen” seems quite false. Prefixing the sentence with “according to the story” does not seem to adequately account for why this sentence is false. It is indeed false that “according to the story, *Hamlet*, Hamlet was created by Jane Austen.” But it is false that Hamlet was created by Jane Austen not because of something indicated by the text of *Hamlet*, but because of certain relations that obtained between the real life entities of William Shakespeare, Jane Austen, and Hamlet.

Sentences such as “Hamlet was created by Shakespeare” must remain unprefixed. If one accepts that “Hamlet was created by Shakespeare” is true, then one must still postulate an object, Hamlet, to account for the truth of this sentence. As useful as the “according to the story” operator can be, it ultimately does not allow fictional anti-realism to avoid postulating fictional objects in every situation.
Given the limited effectiveness of the implicit prefix strategy, fictional anti-realists have offered other ways to account for fictional discourse without postulating fictional entities as objects. A relatively recent anti-realist view that has attracted much attention is known as “pretense theory.” Kendall Walton is the originator of pretense theory and remains its most noted advocate.\(^\text{79}\)

According to Walton’s pretense theory, whenever a person reads a fictional work, reflects upon the work and its characters, and engages in discourse about the work and its characters, the person is engaged in a game of make-believe, in other words a pretense. For example, when one reads *The Adventures of Tom Sawyer*, she temporarily sets aside her belief that the text is a work of fiction, her belief that no boy named Tom Sawyer ever lived in Missouri in the 1850s, and her belief that the event of Tom Sawyer attending his own funeral never happened. Instead she pretends that *The Adventures of Tom Sawyer* is a non-fictional work that accurately describes real people who lived in 1850s Missouri and real events that occurred in this time and place.

Walton uses the term “game of make believe” quite literally. The game of make believe that a reader plays when she reads a fictional work is almost identical the games

---

of make believe that small children play. Two children may choose to play “cops and robbers” in their backyard with nothing more than a tree stump and a stick as props. While the children are playing their game, they must set aside their beliefs that the stump is not a bank teller’s counter, the stick is not a gun, and the children are not a police officer and hardened criminal, respectively. For the purposes of the children’s game, they must pretend that the tree stump is a bank teller’s counter, the stick is a gun, that one child is a police officer, and that the other child is a hardened criminal.

While playing their cops and robbers game, the children will most likely make statements. The children may say things like, “help, I’m being robbed” or “stop, you thief!” The children make these kinds of statements solely within the context of their game. They do not “step outside” the game, as it were, when they say things like “I’m being robbed.” The statement “I’m being robbed” can only be spoken as part of the game. The children make such statements within their game in order to describe facts about and states of affairs within their game, just as real life bank teller would make the statement “I’m being robbed” to describe a real fact about a real life bank robbery.

During the “The Adventures of Tom Sawyer game,” the adult reader will also have the opportunity to make statements. She may say things like “Tom Sawyer lived in Missouri” or “Tom Sawyer attended his own funeral.” But whenever the reader makes statements of this sort, she is still engaged in the game of make believe. She does not “step outside” the game of Tom Sawyer when she makes statements about the novel, its characters, and the properties these characters have. Rather any statement the reader

---

80 Walton, Mimesis as Make Believe, 391-392.
81 Walton, Mimesis as Make Believe, 392.
makes about the people or events found in the world of *Tom Sawyer* can only be made as part of the game.

Thus the kinds of sentences that fictional realists believe require the postulation of fictional objects - “Tom Sawyer attended his own funeral,” “Lady MacBeth is malicious,” “Holmes lives at 221B Baker Street,” and so on - can only be spoken during the course of games of make believe, according to Walton. This leads to some interesting consequences. Walton believes that while people appear to make statements during games of make believe they do not *actually* make statements. They are only pretending to make statements. A child who yells “I’m being robbed” in the course of cops and robbers does not actually assert that he is being robbed; instead he “makes it fictional of himself” that he asserts that he is being robbed.\(^8^2\) It is not just fictional that the child is being robbed, it is also fictional that the child has *asserted that he is being robbed*.\(^8^3\)

Likewise an adult reader of *The Adventures of Tom Sawyer* (Walton refers to her as “Sally”) might say things like “Tom Sawyer attended his own funeral” while she is reading, reflecting upon, and discussing the novel *The Adventures of Tom Sawyer*. When Sally says “Tom Sawyer attended his own funeral” she is still participating in a game of make-believe. Therefore Sally does not actually assert that “Tom Sawyer attended his own funeral.” Instead, Sally *pretends to assert* that, “Tom Sawyer attended his own funeral.” As Walton puts it:

> When Sally says “Tom Sawyer attended his own funeral” in pretense, it is *fictional* in her game that there is someone whom she calls Tom Sawyer and who she claims attended his own funeral. We need not suppose even that there is someone, or some character, about whom Sally *pretends* to speak, nor that there is a proposition she pretends to assert. Her utterance is fully intelligible as an act of

\(^8^2\) Walton, *Mimesis as Make Believe*, 391.

\(^8^3\) Walton, *Mimesis as Make Believe*, 392.
participation if we take it to be *merely* fictional, not true, that her words express a proposition, that they express a proposition about someone whom she calls Tom Sawyer, and that she asserts such a proposition to be true.\(^{84}\)

According to Walton’s pretense theory, sentences like “Tom Sawyer attended his own funeral” can never be actually asserted by anyone. These sentences can only be spoken by people who are engaged in a game of make believe. When a person is engaged in a game of make believe, he or she can only pretend to assert things. This is certainly an odd view and it is unclear how well it coheres with commonsense intuitions regarding discourse about fiction. It is especially unclear how pretense theory can make sense of the notion that certain sentences about fiction have truth values. It is widely believed that the sentence “Tom Sawyer attended his own funeral” is true and the sentence “Tom Sawyer traveled on a spaceship” is false. It is difficult to see how to account for the truth values of sentences that no one can ever assert, only pretend to assert.

Walton does offer an account of the truth value of these sentences that tries to respect commonsense intuitions. It is true that “Tom Sawyer attended his own funeral” and “Tom Sawyer traveled on a spaceship” are sentences that can only be spoken during a game of make-believe, the game of *The Adventures of Tom Sawyer*. Walton points out that there are constraints in this game, just as there are constraints in any game of make-believe. The text of *Tom Sawyer* determines certain facts about the world of the game. The world of the game is limited in how it can be by what the text says. It is true within the world of the game that Tom plays hooky, that Tom gets lost in a cave, and that Tom lives in Missouri because this is what the text says. Likewise it is false within the world of the game that Tom...

\(^{84}\) Walton, *Mimesis as Make Believe*, 396.
of the game that Tom travels on a spaceship, that Tom is a winged horse, and that Tom lives in Denmark because the text does not say these things.  

If Sally states “Tom Sawyer attended his own funeral,” she has, in a sense, asserted something true (or more accurately, she has pretended to assert something true). Sally is engaging in the game of Tom Sawyer when she makes this statement, but given the way the world of Tom Sawyer is, Sally has (fictionally) made a true statement about the world of her game. According to Walton:

What makes Sally’s assertion true . . . is simply the fact that it is fictional in her (authorized) game that she speaks truly. In general, when a participant in a game of make-believe . . . fictionally asserts something by uttering an ordinary statement and in so doing makes a genuine assertion, what she genuinely asserts is true if and only if it is fictional in the game that she speaks truly.

Meanwhile, if Sally states “Tom Sawyer traveled to Neptune on a spaceship” she has, in a sense, asserted something false because her statement contradicts the world of the game she is engaged in.

For Sally to pretend to say of someone she refers to as Tom Sawyer that he traveled by spaceship to Neptune would be inappropriate, unacceptable, for then it would be fictional of the speaker . . . that he or she asserts a falsehood.

Walton recognizes that fictional realists appeal to the apparent truth of sentences like “Tom Sawyer attended his own funeral” in order to postulate fictional objects. He believes that pretense theory renders the postulation of fictional objects completely

85 Walton, Mimesis as Make Believe, 397-400.
86 Walton, Mimesis as Make Believe, 399.
87 Walton, Mimesis as Make Believe, 398.
88 Walton, Mimesis as Make Believe, 390-391.
unnecessary. There is no need for Tom Sawyer to be an object in order to account for the truth of “Tom Sawyer attended his own funeral.”

Fictional entities are obviously not needed for this truth condition to be satisfied. Whether it is satisfied or not in a given case depends on the nature of the work in question . . . . It is fictional in Sally’s authorized game that she speaks truly - because . . . for games authorized for Tom Sawyer are such that, given the words of the text, to speak as Sally does is fictionally to speak truly. No such thing as Tom Sawyer comes into the picture at all.  

At this point, Walton’s pretense theory can be subjected to the same criticism that the implicit prefix strategy was. While these anti-realist theories can make sense of sentences like “Hamlet is a prince” and “Tom Sawyer lived in Missouri” there are many other sentences about fictional entities - “Hamlet was created by Shakespeare” or “Gregor Samsa is a fictional character” - that pose serious challenges to anti-realism. It seems that “Gregor Samsa is a fictional character” is a true sentence. This sentence cannot be prefixed with “according to the story.” Nor can this sentence be accounted for by pretense theory. Pretense theory requires all statements about fictional entities to be made within the context of a game of make believe. When a person reads Kafka’s Metamorphoses, he must pretend for the purposes of the game that the people and events described by the text are real. If the reader states, “Gregor Samsa is a fictional character,” he will violate the rules of the game of Metamorphoses which state that Gregor Samsa is not a fictional character, but rather a flesh and blood human being who turns into a cockroach. Yet it seems obvious that “Gregor Samsa is a fictional character” is a true sentence. How can pretense theory account for this without postulating Gregor Samsa as an object?

---

89 Walton, *Mimesis as Make Believe*, 400.
Walton believes that whenever one reads a work of fiction, he or she engages in a game of make believe. But there are other games of make believe that people can participate in. In particular, Walton thinks that it is possible to construe what I will call “external discourse” (the kind of discourse where people “step outside” fictional works, consider the works from a detached perspective, and make statements like “Samsa is a fictional character” and “Hamlet was created by Shakespeare”) as type of game of make-believe. When a reader plays the game of *Hamlet* he temporarily pretends that entities like Hamlet, Ophelia, and Polonius are real. Likewise, when a person plays the game of external discourse, he temporarily pretends that entities like “fictional characters” are real, even though fictional characters are not to be found in the real world.

There may be an unofficial game in which one who says [Gregor Samsa is a fictional character] is fictionally speaking the truth, a game in which it is fictional that there are two kinds of people: “real” people and “fictional characters.” The speaker can be regarded as participating in this unofficial game, as fictionally referring to something by means of the name “Gregor Samsa” and attributing to the referent a property he specifies by the predicate “is a fictional character.”

Because statements like “Gregor Samsa is a fictional character” still are spoken during a game of make-believe, such sentences do not entail fictional realism. For “it does not matter that there is no such property as that of being a fictional character. The speaker is pretending that these words pick out a property; it is fictional in the unofficial game that they do.”

---

90 Walton, *Mimesis as Make Believe*, 423.
Thus concludes my discussion of fictional realism and anti-realism. I am a fictional realist myself. For the remainder of this dissertation I will assume that fictional realism is correct, and I will discuss some of the very interesting consequences that follow if one assumes fictional realism is true. It is beyond the scope of my project to provide a detailed defense of fictional realism. But allow me to provide a very brief explanation of why I find fictional realism to be more plausible than anti-realism.

Like many other philosophers, I find the linguistic evidence in favor of fictional realism very compelling. I share the intuitions of most people that sentences like “Hamlet is a prince” and “Lady MacBeth is a winged horse” have clear-cut truth values. I do not think it is a worthwhile tradeoff to ignore commonsense intuitions about these types of sentences simply for the sake of avoiding ontological commitments to fictional objects.

The fictional anti-realist responses to the linguistic case for realism are not very convincing. The strategy of implicitly prefixing every sentence about fiction with “according to the story” is only effective when applied to a subset of sentences about fiction. Meanwhile I must confess that I find Walton’s pretense theory very puzzling. Walton’s insight that our engagement with fictional texts is akin to a game of make-believe is not completely implausible. It could be the case that when ordinary people read fictional works, they are temporarily suspending belief and pretending that the characters in the work are real people and the world really is the way it is described in the book. Certainly some people engage with fiction in this manner. But I believe Walton
seriously misconstrues external discourse about fiction when he describes it as another game of make-believe. This is a highly implausible account of what most people take themselves to be doing when they engage in external discourse about fiction.

There are other reasons why I am a fictional realist; these reasons will become clearer in the following chapter. I believe fictional realism is more plausible when one provides a thorough account of what kind of objects fictional entities would be before one postulates fictional entities. For there are several different types of fictional realism, all of which make different claims about the nature of fictional entities. Depending on the brand of fictional realism one endorses, fictional entities will turn out to be metaphysically messy objects or logically coherent, philosophically respectable objects that deserve a home in our ontology.
Varieties of Fictional Realism

In the previous chapter, I introduced fictional realism – the claim that fictional entities are *bona fide* objects that deserve a place in our ontology. I considered the various reasons why both historic and contemporary philosophers have adopted fictional realism; I also considered anti-realist arguments against fictional realism. Fictional realism is a plausible view, I concluded, but there is a caveat to this. Depending on what form of fictional realism one adopts – that is, depending on what kind of object one takes fictional entities to be – fictional realism can come off as wildly implausible or metaphysically respectable.

Chapter 2 is a thorough study of several different and mutually exclusive varieties of fictional realism. Meinong’s theory of fiction will be considered in greater detail than it was in the previous chapter. I will also discuss the neo-Meinongian fictional realism of Terence Parsons. A view quite similar to Meinongian fictional realism, Platonist fictional realism, will receive close attention as well. I will then carefully examine a form of fictional realism – idealism – that neatly contrasts with Meinongianism, neo-Meinongianism, and Platonism. After considering the numerous respective drawbacks to these various theories of realism, I will introduce my preferred theory of fictional realism, artifactualism. Artifactualism, as it turns out, best respects our ordinary intuitions and
professional practices surrounding fictional entities; it also avoids the many ontological problems that befall other forms of fictional realism.

2.1 What a Theory of Fictional Realism Ought to Say

All theories of fictional realism share the broad claim that fictional entities are objects. If we assume that fictional entities are indeed objects, the question then becomes what kind of objects are they? All theories of fictional realism must answer this question. All theories of fictional realism should indicate whether fictional entities are existent objects or whether they are non-existential objects. All theories of fictional realism should also take a position on whether fictional entities are concrete spatio-temporal objects, abstract objects, mental objects, or some other kind of object that does not fit neatly into any traditional metaphysical category.

Ideally all theories of fictional realism would include additional information about the nature of fictional entities. A theory should provide criteria for determining which properties a fictional entity has and lacks. Ordinary discourse about fictional entities takes for granted that fictional entities have, in some sense, all of the properties attributed to them by their stories. The story *Hamlet* explicitly predicates the property being a prince of Hamlet; therefore Hamlet has the property of being a prince. But this matter is actually rather complicated. When one says Hamlet has the property of being a prince, does she mean to say that Hamlet has the actual property of being a prince just as Charles Windsor has this property? Or is her predicative statement implicitly prefixed with “according to the story,” with the result that Hamlet does not actually have the property
of being a prince (as Charles Windsor does) but rather has the property of being a prince according to the story, this latter property being an entirely different one?

Another question is whether or not fictional entities possess properties that are implied but not explicitly predicated by their stories. For instance, to the best of my knowledge, *Pride and Prejudice* never explicitly states that Lydia Bennett has the property of having feet. Any minimally educated reading of the novel seems to imply that Lydia has feet; she enjoys dancing, she climbs stairs, and there is nothing in the text that indicates she has the striking feature of lacking feet. Still it is important for a theory of fictional realism to indicate whether or not we can attribute implicitly implied properties to fictional entities or if fictional entities are restricted to only the properties explicitly predicated of them in their stories.

The impossibility of some fictional entities and incompleteness of all fictional entities are central to anti-realism about fiction. The short “stories” of Anthony Everett provide several cases of fictional entities that blatantly violate laws of logic, like symmetry of identity.\(^{92}\) Meanwhile every fictional character appears to be radically incomplete; the number of Lady MacBeth’s children being a well-known case. Ideally any theory of fictional realism will explain how fictional entities can seemingly violate the law of non-contradiction and be incomplete while still remaining *bona fide* objects. But as we will see, not all the forms of fictional realism considered in this chapter address this problem.

Another issue that ideally ought to be addressed by all theories of fictional realism is the status of fictional works. Clearly fictional entities have a very important relationship with their stories. Already I have mentioned the vital role a fictional work

\(^{92}\) Everett, “Against Fictional Realism,” 634.
has in determining the properties of fictional entities. If a work of fiction ascribes a certain property to a fictional entity, then that fictional entity seems to have that property. Hamlet has the property of being a prince because *Hamlet* explicitly ascribes that property to him. Hamlet very well may have other properties that *Hamlet* does not explicitly ascribe to him (such as the property of being a fictional character, or the property of having an Oedipus complex). Those are more controversial matters which will be discussed in more detail later.

Works of fiction also play an important role in our ordinary, pre-theoretic beliefs about fictional entities. It is very common to speak of fictional characters as being “in” works of fiction. I think that everyday discourse about fictional entities being “in” works of fiction belies a widespread intuition that fictional entities are in some sense dependent on fictional works in order to exist.\(^{93}\) While it is always difficult to pin down folk intuitions about philosophical topics, I do not think it is rash to suppose that most people think that characters such as Hamlet and Ophelia would not exist if *Hamlet* had never been written.

Since fictional works play a vital role in determining some of the properties of fictional entities, and perhaps also play a role in bringing fictional entities into existence or keeping them in existence, it would be good for any theory of fictional realism to say something about fictional works. In particular, it would be very interesting if a theory of fictional realism said whether or not fictional works and fictional entities are in the same ontological category. If fictional entities are, for example, Platonic objects, are fictional works Platonic objects as well?

---

\(^{93}\) Or, to put it in Meinongian terms, fictional entities are in some sense dependent on fictional works in order to be objects.
Unfortunately not every form of fictional realism discussed in this chapter has something interesting to say about fictional works. This may be due to the fact that realism about fictional entities is enormously controversial while, to the best of my knowledge, there is no corresponding controversy over realism about fictional works.\textsuperscript{94} Much philosophical energy has been devoted to the question of whether or not Hamlet and Elizabeth Bennett are objects, but no one seems to deny that \textit{Hamlet} and \textit{Pride and Prejudice} are.\textsuperscript{95}

As we consider various theories of fictional realism, we will see inconsistent treatment of fictional works. Some philosophers, such as Parsons, have a great deal to say about the ontology of fictional entities, but nothing to say about the ontology of fictional works. Other philosophers such as Benedetto Croce and R.G. Collingwood focus almost entirely on the ontology of fictional works to the exclusion of fictional entities. At times a fair amount of speculation must be used to draw out their views on the ontology of fictional entities. Meanwhile Nicholas Wolterstorff and Amie Thomasson give accounts of the ontology of both fictional works and fictional entities.

Finally any solid theory of fictional realism ought to give an account of the role of the author and his or her relationship to fictional entities. Obviously the author has a very important relationship to fictional entities. Ordinary intuition holds that authors create fictional works and fictional entities; it common to hear authors compared to gods or parents. It also seems fair to say that authors determine what their fictional entities are like - they decide which properties the characters will have.

\textsuperscript{94} Thomasson, \textit{Fiction and Metaphysics}, 141, 143, 148.
\textsuperscript{95} This is not to say that there is no controversy over what \textit{kind of object} literary works are.
But depending on what kind of object fictional works and entities turn out to be, the intuitive picture of authors as creators may be quite false. Any theory of fictional realism ought to indicate whether or not our pre-theoretic intuitions about the relationship between authors and fictional entities are true. Whether a theory of fictional realism ultimately concurs with or rejects folk intuitions about the role of the author, the theory should give a thorough account of the role of the author in bringing fictional entities into existence and determining what properties fictional entities have. I cannot overstate how important it is for all theories of fictional realism to have a solid account of the relationship between the author and fictional entities.

2.2 Meinongian Fictional Realism

Since contemporary fictional realism has its origins in the metaphysics of Meinong, I shall begin with an account of Meinongian fictional realism. By Meinongian fictional realism, I mean any theory of fictional realism that takes fictional entities to be Meinongian non-existent objects.96

Discovering the views held by Meinong himself about the nature of fictional entities can be difficult. Meinong wrote relatively little about fictional entities and aesthetics in general.97 This is surprising given that Meinong personally had a great

96 It should be noted that Meinongianism itself is not necessarily committed to Meinongian fictional realism. A philosopher could embrace Meinong’s overall theory of objects and accept that there are non-existent objects which, despite lacking being entirely, still are objects with properties, yet she could also hold that fictional entities are not these kinds of objects. Rather, she could argue, there are good reasons to think fictional entities ought to be placed in the Meinongian category of existent objects. I am very sympathetic to this view myself.

interest in the arts, especially music. One might expect Meinong to have written at least one paper solely devoted to the ontology of music or of art. But instead, Meinong’s views on aesthetics and specifically on fiction are found scattered throughout his other writings expounding on the views he put forth in *The Theory of Objects*.

Meinong first had an opportunity to apply his overall ontology to fictional entities when he responded to Theodor Lipps’ essay “Further Considerations on Empathy.” Lipps had argued that fictional entities, such as Goethe’s Mephistus, are “called into existence” by their authors. Once a fictional entity is called into existence by its author, the entity possesses a type of reality which Lipps termed “aesthetic reality.” Meinong considered and rejected Lipps’ views in the 1905 essay, “Judgment-feelings: what they are and what they are not.” Meinong argued that it is erroneous to claim that there is a special type of reality, aesthetic reality, somehow distinct from ordinary reality. There is only one kind of reality, and it is only possessed by concrete, spatio-temporal objects. It is important to remember that for Meinong, terms like “reality” or “real” are synonymous with existence. The only objects which exist and are thus “real” are concrete, spatio-temporal objects. If an object is not spatio-temporal, then it is non-existent and unreal. Of course, non-existent unreal objects still are objects and still possess properties.

Even though Lipps erred in setting up “aesthetic reality” as distinct from ordinary reality, was he still correct that fictional entities like Mephistus are real objects brought into existence by their authors? No, answers Meinong. Fictional entities are unreal, non-

---

98 Schuhmann, “Meinongian Aesthetics,” 517.
99 In the original German, “Weiteres zur Eingehlung,” see Raspa, “Fictional and Aesthetic Objects,” 52. This essay has only been partially translated into English. All quotes from Lipps are of Raspa’s translation.
100 Raspa, “Fictional and Aesthetic Objects,” 52-53.
101 In the original German, “Über Urteilsgefühle: was sie sind und was sie nicht sind,” see Raspa, “Fictional and Aesthetic Objects,” 52. All quotes from Meinong are of Raspa’s translation.
102 Raspa, “Fictional and Aesthetic Objects,” 49.
existent objects. No author can bring a fictional entity into existence, as Lipps claims, because fictional entities do not and cannot exist in the first place.\textsuperscript{103} Why are fictional entities disqualified from existence? According to Meinong, it is because fictional entities lack spatio-temporal properties entirely. A fictional entity like Mephistus “does not belong to the past, but to the immediate present.”\textsuperscript{104} Confusingly Meinong uses the term “the immediate present” to denote a kind of timelessness. Elsewhere Meinong speaks of eternally and analytically true propositions like “the equilateral triangle has equal angles” as belonging to the immediate present.\textsuperscript{105} Thus when Meinong claims that Mephistus belongs to the immediate present, he really means that Mephistus is a timeless entity. The same applies to all fictional entities. Fictional “characters lack the past and also the future.”\textsuperscript{106}

It should be noted that Meinong did not believe that fictional entities were unique among art works in their non-existence. Other forms of art such as literary works, music, and paintings – which Meinong referred to using the umbrella term “the aesthetic object”\textsuperscript{107} – were similarly outside of time and thus non-existent.\textsuperscript{108} An aesthetic object such as Pride and Prejudice or Beethoven’s Ninth Symphony is just one mere object “among the infinite totality of extra-ontic objects.”\textsuperscript{109}

By now we are quite familiar with Meinong’s view that non-existent objects are still objects and possess properties in spite of their non-existence. Fictional entities, like all other non-existent objects, possess properties. But as I noted in the previous section,

\textsuperscript{103} This point is also noted by Thomasson in Fiction and Metaphysics 15-17.
\textsuperscript{104} Meinong, quoted in Raspa, “Fictional and Aesthetic Objects,” 54.
\textsuperscript{105} Raspa, “Fictional and Aesthetic Objects,” 54-55.
\textsuperscript{106} Meinong, quoted in Raspa, “Fictional and Aesthetic Objects,” 55.
\textsuperscript{107} Raspa, “Fictional and Aesthetic Objects,” 47.
\textsuperscript{109} Meinong, quoted in Schuhmann, “Meinongian Aesthetics,” 534.
there is much philosophical complexity and controversy beneath the simple statement “fictional entities have properties.” Unfortunately for our purposes, Meinong’s take on which properties fictional entities have and lack is quite undeveloped. In his 1917 essay “On Emotional Presentation,” Meinong adopts the uncontroversial position that if a literary work explicitly ascribes a property to a fictional entity then that fictional entity truly has that property.\textsuperscript{110}

In modern drama, exact information is frequently given as to the age or other properties of the characters in the play. At first this information can only be of the order of assumptions. But once such assumptions are made, the characters in question \textit{are} indeed of the indicated age.\textsuperscript{111}

But beyond this, Meinong does not provide answers to the questions that ideally a theory of fictional realism ought to answer. Meinong takes no position on whether a fictional entity has properties in addition to those explicitly stated in the text; for instance, does Lydia Bennett have the property of having feet, even though \textit{Pride and Prejudice} never explicitly ascribes this property to her? Meinong takes no position on the issue of how some fictional entities can seemingly violate the law of non-contradiction and how all fictional entities are incomplete.

Given the silence in Meinong’s writings, we cannot know if he ever noticed that fictional entities frequently appear to violate the laws of non-contradiction and always are incomplete. But it is reasonable to suppose that had Meinong noticed this fact, it would not have bothered him much, and it certainly would not have led him to reject fictional realism. Meinong’s ontology already embraces impossible objects such as the round

\textsuperscript{110} Raspa, “Fictional and Aesthetic Objects,” 58-59.
\textsuperscript{111} Meinong, quoted in Raspa, “Fictional and Aesthetic Objects,” 59.
square.\textsuperscript{112} Under Meinong’s ontology, the status of fictional entities would not change in any meaningful way if it turned out that fictional entities were actually impossible objects. If Hamlet violates of the law of non-contradiction or is incomplete, then he is an impossible object, and thus a non-existent object. If Hamlet does not violate the law of non-contradiction and if he is complete, he is still a fictional entity, and thus a non-existent object.

Ideally a theory of fictional realism should have something to say about fictional works, since there is obviously an important relationship between fictional entities and the fictional works in which they appear. Meinong, as we have already seen, classifies fictional works like plays and novels as “aesthetic objects.” All aesthetic objects are non-existent objects.\textsuperscript{113} Therefore, \textit{Hamlet}, \textit{Pride and Prejudice}, and the like are non-existent objects. This is a strange-sounding claim to be sure, especially because it entails that \textit{Hamlet} and \textit{Pride and Prejudice} are not spatio-temporal entities. Surely one might argue, \textit{Pride and Prejudice} is a spatio-temporal object - after all, it is presently sitting on my bookshelf, it is about seven inches tall, weighs a little less than a pound, and is made of paper. These are all spatial properties.\textsuperscript{114} Furthermore \textit{Pride and Prejudice} seems to have temporal properties. My personal copy of it is less than ten years old, and the novel itself came into existence in the early nineteenth century.

Meinong would not deny that my personal copy of \textit{Pride and Prejudice} is an existent, spatio-temporal object. But he would deny that my personal copy of \textit{Pride and Prejudice} or any other copy of it in existence is identical to the work, \textit{Pride and Prejudice}.

\textsuperscript{112} Meinong, “The Theory of Objects,” 82.
\textsuperscript{113} Raspa, “ Fictional and Aesthetic Objects,” 56; Schuhmann, “Meinongian Aesthetics,” 534.
\textsuperscript{114} This is only a hypothetical counter-argument; it is not my own position.
Prejudice, itself. Meinong makes a distinction, quite common in aesthetics, between a work of art and its instances or copies. While a printed copy of a literary work is not the literary work itself, a printed copy of a book (or a dramatic script or a musical score) stands in an important relationship to the aesthetic object itself. The printed copy of Pride and Prejudice “signifies” and “points towards” Pride and Prejudice itself.

Through the physical book, the reader discovers Pride and Prejudice, apprehends its nature of the work, and apprehends the characters within it. This is no trivial feat, given that a reader is a concrete, existing, object and a literary work is a non-existent, abstract, object. Physical things like books, scripts, or musical scores, bridge the divide between spatio-temporal readers and non-spatio-temporal, non-existent, aesthetic objects.

Finally let us turn to Meinong’s views about the role of the author. Ordinary intuition indicates that an author brings fictional entities into existence. Meinong’s theory of fictional realism rejects this intuition for two reasons. First, fictional entities are non-existent objects; it makes no sense to speak of bringing a non-existent object into existence. Secondly and more importantly, as Meinongian non-existent objects, fictional entities (as well as all other kinds of aesthetic objects) are eternal and timeless objects that cannot be created or destroyed. Despite commonsense intuitions about the role of the author, “the aesthetic object as such neither exists nor can it be created.”

Not surprisingly, Meinongian fictional realism also rejects the folk intuition that the author determines some of the properties of his fictional entities. When reading Faust one notes that Mephistus has various properties, such as being a devil and making a pact.

---

117 Raspa, “Fictional and Aesthetic Objects,” 58.
118 Meinong, quoted in Schuhmann, “Meinongian Aesthetics,” 534.
with Faust. Ordinarily one would think that Mephistus has these properties because these are the properties that Mephistus’ author, Goethe, chose to give to him. Had Goethe decided to give Mephistus different properties, such as the property of making a pact with Gretchen, then Mephistus indeed would have had that property. Ultimately, folk intuition holds, it is “up to” the author which properties a fictional character has and lacks.

But under Meinong’s theory of fictional realism, authors like Goethe have no such creative power over the properties of their fictional entities. Because Mephistus is an eternal, non-existent object, all of his properties are pre-determined and immutable.119 Centuries before Goethe was born, there was an aesthetic object, Mephistus, with the property of striking a deal with Faust. Goethe in no way created Mephistus or gave him that property. Rather when Goethe wrote Faust he “picked out” an aesthetic object, Mephistus, “among the infinite totality of extra-ontic objects.”120 Goethe was more akin to a selector, rather than creator, of Mephistus. And when Goethe wrote of how Mephistus made a pact with Faust, he was not exercising any creative power of his own and attributing a new property to an object that did not yet have it and could have lacked it. Instead Goethe was reporting an analytic judgment about an object, Mephistus, that already had the property of making a pact with Faust.121

---

119 Raspa, “Fictional and Aesthetic Objects,” 58, 60.
120 Meinong, quoted in Schuhmann, “Meinongian Aesthetics,” 534.
2.3 Neo-Meinongian Fictional Realism

Meinongianism has enjoyed a modest revival in recent decades. While remaining faithful to the core ideas of Meinong, neo-Meinongianism adds new details in an attempt to deal with the apparent inconsistencies and contradictions in Meinong’s original theory. Of the many contemporary neo-Meinongians, none have provided as complete an account of fictional realism as Terence Parsons has. His neo-Meinongian theory of fictional realism will be the sole focus of this section.

Parsons’ motivation for fictional realism is quite different from Meinong’s. Meinong’s motivation for fictional realism was his overall theory of objects which was, in turn, motivated by Brentano’s theory of intentionality. But Parsons embraces fictional realism because of what I have called the linguistic case for fictional realism. In *Nonexistent Objects*, Parsons notes that there are many propositions that strike us as intuitively true or false which refer to “unreal objects,” including mythological beings like the Greek gods and fictional entities like Sherlock Holmes. In order to account for our intuition that a sentence like “Ironically, a certain fictional detective (namely, Sherlock Holmes) is much more famous than any real detective, living or dead” is true, Parsons argues that we must accept that Sherlock Holmes is an object. Parsons also points out that we can construct valid inferences out of sentences that refer to fictional

---


123 Parsons, *Nonexistent Objects*, 32.
entities. The only way to account for this fact is to postulate fictional entities as objects.\textsuperscript{124}

Parsons accepts Meinong’s characterization of fictional entities as a type of non-existent, unreal, object.\textsuperscript{125} But as a neo-Meinongian, Parsons takes a different approach to unreal objects, particularly regarding the kinds of properties unreal objects have.

Before focusing further on Parsons’ views on fictional entities, we need to discuss his overall theory of predication.

All objects, real and unreal, have properties; this is standard Meinongian dogma. Parsons adds a twist by claiming that all properties possessed by objects fall into one of two categories: nuclear properties and extranuclear properties (sometimes referred to as nuclear and extranuclear predicates). Nuclear properties tend to be “ordinary” properties such as being blue, being tall, kicking Socrates, being kicked by Socrates, and being a mountain.\textsuperscript{126} Extranuclear properties, on the other hand, tend to be more exotic. They are “ontological” properties like being existent, being mythical, and being fictional. “Modal” properties like being possible and being impossible, “intentional” properties like being thought about by Meinong and being worshipped by someone, and “technical” properties like being complete, are also extranuclear properties.\textsuperscript{127}

There is no strict decision procedure for determining whether a property is nuclear or extranuclear. Parsons believes that the best one can do is provide many examples of properties that are clearly nuclear or clearly extranuclear in hopes that we “develop an

\textsuperscript{124} Parsons, \textit{Nonexistent Objects}, 34.
\textsuperscript{125} Parsons, “Are there Nonexistent Objects?” 366; \textit{Nonexistent Objects}, 32-33, 49-51, 54; “A Prolegomenon,” 566.
\textsuperscript{126} Parsons, \textit{Nonexistent Objects}, 23; “A Prolegomenon,” 569-570.
\textsuperscript{127} Parsons, \textit{Nonexistent Objects}; “Nuclear and Extranuclear Properties”; “A Prolegomenon,” 569-571.
intuition for the distinction, so that we can readily classify new cases.”

Parsons notes that he has “such an intuitive ability, and that other people pick it up quite readily; even those who are skeptical about the viability of the distinction seem to agree about which predicates are supposed to be which.”

Every real object has both nuclear and extranuclear properties; every unreal, non-existent object also has both nuclear and extranuclear properties. Mount Everest has nuclear properties like being a mountain, being 29,029 feet high, and being in Nepal, and has extranuclear properties like being existent, and being possible. Meinong’s famous golden mountain has nuclear properties like being a mountain, being made of gold, and extranuclear properties like being non-existent, being possible, and being thought about by Meinong. Lydia Bennett has nuclear properties like being female, and being flirtatious, and extranuclear properties like being non-existent and being fictional.

Parsons believes that the distinction between nuclear and extranuclear properties originated with Meinong, or at least was what Meinong had in mind. The distinction was further developed in the work of one of Meinong’s students, Ernst Mally. Later neo-Meinongians such as Richard Routley and Hector Castaneda also make use of the distinction between nuclear and extranuclear properties.

Parsons avails himself of the nuclear and extranuclear property distinction to avoid falling prey to the devastating critique Bertrand Russell made of Meinong’s

---

128 Parsons, Nonexistent Objects, 24.
129 Parsons, Nonexistent Objects, 24.
132 Castaneda 11-24; Parsons, “A Prolegomenon,” 561; Routley 537-605.
original theory, discussed in the previous chapter. Russell noted that Meinong’s theory of objects “regards any grammatically correct denoting phrase as standing for an object. Thus ‘the present King of France,’ ‘the round square,’ etc, are supposed to be genuine objects.” Russell then pointed out that given Meinong’s theory, there must also be an object denoted by the phrase “the existent King of France.” Meinong and Russell both would agree that the existent King of France clearly does not exist. Unfortunately for Meinong, this leads to a violation of the law of non-contradiction. The existent King of France has the property of being existent and the property of being non-existent. Russell claimed that this result was “intolerable” and was sufficient grounds for rejecting Meinong’s entire theory of objects.

Meinong never provided a convincing response to Russell’s critique, and the reputation of Meinong’s theory of objects suffered greatly because of this. But Parsons recognizes that he must provide a response to the Russellian objection, if neo-Meinongianism is to be tenable. According to Parsons’ neo-Meinongianism, there is an object correlated with every non-empty set of properties. There is a non-empty set of properties that includes the properties of being female, being a scientist, being named Marie Curie, and the like, and there is an object correlated with this set - Marie Curie. There is a non-empty set of properties that includes the properties of being golden, being a mountain, and being non-existent. The object correlated with this set of properties is the famous golden mountain. But there is also a non-empty set of properties that consists

---

136 Parsons, Nonexistent Objects, 4; “A Prolegomenon,” 562-563.
137 Parsons, Nonexistent Objects, 17-19.
of the properties of being a King, being the King of France, and being existent. The object correlated with this set of properties is the troublesome existent King of France.

The distinction between extranuclear and nuclear properties solves the problem of the existent King of France. It is not actually the case, Parsons says, that there is an object correlated with every non-empty set of properties. This statement is too broad. Rather, there is an object correlated with every non-empty set of nuclear properties.\textsuperscript{138}

We do not have to worry about contradictory objects like the existent (yet non-existent) King of France. There is no object correlated with the set of properties: being a King, being the King of France, being existent, and being non-existent. No object can be correlated with this set of properties because two of the properties in this set, being existent and being non-existent, are extranuclear properties.

The nuclear/extranuclear property distinction is put to great use in Parsons’ theory of fictional realism. First let us consider what properties fictional entities have, according to Parsons’ theory. A fictional entity like Hamlet obviously has the property of being fictional. That is an extranuclear property. As for his nuclear properties, Hamlet has “all and only” the nuclear properties attributed to him by his story.\textsuperscript{139} Since according to \textit{Hamlet}, Hamlet has nuclear properties like being Danish, being a prince, and being the son of Gertrude, Hamlet the fictional entity really does have those properties.\textsuperscript{140}

Earlier I noted that fictional entities seem to possess properties that are never explicitly mentioned by their stories. For example, it seems quite reasonable to suppose that Hamlet has the nuclear property of having feet, despite the fact that the text of

\textsuperscript{138} Parsons, \textit{Nonexistent Objects}, 19.
\textsuperscript{139} Parsons, \textit{Nonexistent Objects}, 183.
\textsuperscript{140} Note that such predicative statements are not to be implicitly prefixed with according to the story. Hamlet has the property of being a prince, not the property of being a prince \textit{according to the story}. 
Hamlet never explicitly mentions this. Parsons believes that fictional entities do indeed have nuclear properties that, while not explicitly stated by their texts, are implied by a minimally educated reading of their texts.

Any reader who failed to extrapolate far beyond the printed word would be quite unable to understand the story. This is typically not a conscious process; we just instinctively assume all sorts of things on the basis of the given information, just as we do in perception. We read “Commander Roderick Blaine looked frantically around the bridge . . .” and we immediately assume, without conscious reasoning, that Blaine is named ‘Blaine’, that he is a commander, that he is male, that his eyes are open, that he is located in a space more than one cubic inch in volume, that he is upset . . . Any of these inferences may be revised in the light of further information (e.g., if we read that his eyelids had been replaced with some transparent material, and were now fastened shut).

Fictional works not only attribute nuclear properties to fictional entities, they also attribute extranuclear properties to fictional entities. According to Hamlet, Hamlet, Ophelia, and Polonius all have the extranuclear property of being existent. This is a problem for neo-Meinongianism, given that fictional entities are taken to be non-existent objects. If fictional entities have the extranuclear properties attributed to them by their stories, just as they have the nuclear properties attributed to them by their stories, then Hamlet becomes a contradictory object quite similar to the existent King of France. Hamlet is both existent (by virtue of having that property in Hamlet) and non-existent (by virtue of being a fictional entity). To dodge this unappealing scenario, Parsons is forced to conclude that fictional characters do not, as a general rule, have the extranuclear properties attributed to them by their stories.

141 Parsons, Nonexistent Objects, 176-177.
Typically, stories attribute to their characters extranuclear properties as well as nuclear ones. The theory in question says that these characters have the nuclear properties so attributed, but does not say this for the extranuclear ones. In many cases this is essential. According to the story, Holmes exists; but we would not want to attribute existence to him on that account, for he does not exist.  

Occasionally a fictional entity will actually have an extranuclear property that is also ascribed to it by the story. As an example of this, Parsons offers Sherlock Holmes, who according to the Sherlock Holmes stories, has the extranuclear property of being thought about by criminals. Outside of the world of the stories, Sherlock Holmes also has the extranuclear property of being thought about by criminals. But this is merely a coincidence. Sherlock Holmes only possesses that extranuclear property because some criminals have read Conan Doyle’s stories, not because the stories ascribe this property to Holmes.  

Opponents of fictional realism note that fictional entities routinely violate the laws of non-contradiction and excluded middle. Since no object can violate these laws, fictional entities cannot be objects and fictional realism must be false. Parsons, like Meinong, is unfazed by this objection. He readily admits that some fictional entities violate the law of non-contradiction by possessing incompatible nuclear properties within their stories. All fictional entities violate the law of excluded middle due to the fact that no fictional work is lengthy and detailed enough to explicitly state whether a given fictional entity has or lacks every nuclear property. “Characters native to (created in) a story have all and only those nuclear properties attributed to them in the story. This

142 Parsons, Nonexistent Objects, 197-198.  
143 Parsons, Nonexistent Objects, 198.  
144 Parsons, Nonexistent Objects, 183-184.
means that such objects will typically be highly incomplete, and will occasionally be impossible.”

The anti-realist assertion that no object can violate the laws of non-contradiction and excluded middle, is only half-correct under neo-Meinongianism. It is more accurate to say that no *existent* object can be impossible or incomplete. Non-existent objects can be impossible, and non-existent objects are usually incomplete. Fictional entities, being non-existent, can routinely flout the laws of non-contradiction and excluded middle with no damage to their status as objects.

Ideally a theory of fictional realism will have something to say about fictional works. Parsons never discusses fictional works and their ontological status. It remains a mystery whether Parsons holds Meinong’s view that aesthetic objects like literary works are non-existent, or whether he believes that literary works exist and are not particularly controversial compared to fictional entities.

As for the relationship between fictional entities and their authors, Parsons must follow Meinong’s path and reject the intuitive beliefs that authors bring fictional entities into existence and determine the characteristics of those entities. Like Meinongianism, neo-Meinongianism holds fictional entities to be non-existent objects, and it is nonsensical to speak of a non-existent object being brought into existence.

I have said that, in a popular sense, an author *creates* characters, but this too is hard to analyze. It does not mean, for example, that the author brings those characters into existence, for they do not exist. Nor does he or she make them objects, for they were objects before they appeared in stories. We might say, I suppose, that the author makes them *fictional* objects and that they were not fictional objects before the

---

146 Parsons does note that some nonexistent objects are complete, see *Nonexistent Objects*, 20.
creative act. We might even say that the author bestows on them fictional existence . . . as long as this is not confused with ordinary existence.\textsuperscript{147}

For Parsons, an author’s creative power is nothing like a god’s or a parent’s. An author can “pick out” a non-existent object and write about the object in a story. Due to the creative actions of the author, this non-existent object will have a property, being fictional, that it did not previously have. But as Parsons points out, this is not the same thing as creating an object \textit{ex nihilo}.

\subsection*{2.4 Platonist Fictional Realism}

1980 was a fruitful year for fictional realism, seeing the publications of \textit{Nonexistent Objects} and Nicholas Wolterstorff’s \textit{Works and Worlds of Art}. Wolterstorff puts forth a Platonist theory of fictional realism.\textsuperscript{148} A Platonist theory of fictional realism is characterized by the claims that fictional entities are objects, these object exist (a crucial difference with Meinongianism) and these objects, like Platonic forms, exist eternally and necessarily. Wolterstorff believes that fictional entities, as well as literary works and some other art forms, are best viewed as natural kinds.

This interesting conclusion is reached via a general examination of the nature of art. Wolterstorff notes that when one considers art in general one notices an interesting fact. Certain types of art, particularly music, drama, and literature, have many instances

\textsuperscript{147} Parsons, \textit{Nonexistent Objects}, 188.
\textsuperscript{148} Wolterstorff himself does not call his view “Platonist.” Instead other philosophers have termed it such. See Everett, “Against Fictional Realism,” 626 and van Inwagen, “Existence, Ontological Commitment, and Fictional Entities,” 150 n. 19, 153. For the sake of simplicity of discourse, I will also call this theory “Platonist fictional realism” as opposed to “fictional entities as natural kinds form of fictional realism” or “Wolterstorff’s fictional realism” even though Wolterstorff, as far as I can tell, is the only philosopher who advocates this particular form of fictional realism.
or examples, which are distinct from the work of art itself. Or consider, for instance, the musical work Verdi’s *Requiem*. Like all great compositions, this one musical work has been performed multiple times: in Italy at its late nineteenth century premiere, at the University of Notre Dame in February of 2001, and countless other times. The 2001 performance of *Requiem* and the premiere performance of *Requiem* are both examples of *Requiem*. But no performance of *Requiem* is identical to *Requiem* itself. While every performance of *Requiem* shares some properties with the work *Requiem* (such as being composed by Verdi), every performance of *Requiem* will also have properties that *Requiem* the work does not have. One performance might be overly fast-paced, plodding, off-key, or technically perfect. However *Requiem simpliciter* clearly does not have properties like being plodding, off-key, or technically perfect. Thus there is a distinction between works of art, like Verdi’s *Requiem*, and their examples.

The same applies to art works like drama and literature. *Hamlet* is a dramatic work; all of the different performances of *Hamlet* over the centuries are examples of *Hamlet*, distinct from *Hamlet* itself. Nikolai Gogol’s *Dead Souls* is a literary work; all of the different physical copies of *Dead Souls* as well as all of the different “utterances” or readings of *Dead Souls* are examples of this literary work, distinct from *Dead Souls simpliciter*.

---

151 Wolterstorff, *Works and Worlds of Art*, 38-39. Note that Wolterstorff does not take printed scripts of a drama to be examples of the drama, but rather performances.
Wolterstorff points out that the relationship between an artwork and its examples, particularly in how properties are shared between an artwork and its examples, is remarkably similar to the relationship between natural kinds and their examples.\textsuperscript{153} This is more than mere coincidence. Wolterstorff argues that artworks are ontologically analogous to natural kinds.\textsuperscript{154} Works of art, including fictional works like \textit{Hamlet} and \textit{Dead Souls}, are properly viewed as natural kinds.\textsuperscript{155} Natural kinds exist from eternity and exist necessarily.\textsuperscript{156} For this reason, Wolterstorff’s theory of fictional realism is best termed “Platonist.”

According to Wolterstorff both fictional works and fictional entities are natural kinds. Sherlock Holmes, Hamlet, and Ophelia are all natural kinds. The natural kind that is a fictional entity always refers to the fictional work that fictional entity is in. In other words, the natural kind that is Ophelia is properly termed the Ophelia-in-\textit{Hamlet} kind. The natural kind that is Hamlet is properly termed the Hamlet-in-\textit{Hamlet} kind. Chichikov, the protagonist of \textit{Dead Souls}, is “a certain kind of person, that is, not a person of a certain kind, but a certain person-kind - the Chichikov-in-\textit{Dead-Souls} kind.”\textsuperscript{157}

Sometimes one and the same fictional character appears in two different fictional works, or so it seems. For example, the fictional entity Faust appears in Marlowe’s \textit{Dr. Faustus} as well as in Goethe’s \textit{Faust}. In this case, Wolterstorff says, we must distinguish between two natural kinds the Faust-in-Marlowe’s-\textit{Dr. Faustus} kind and the Faust-in-\textit{Dr. Faustus} kind.

\begin{thebibliography}{9}
\bibitem{153} Wolterstorff, \textit{Works and Worlds of Art}, 46.
\bibitem{154} Wolterstorff, \textit{Works and Worlds of Art}, 46-62.
\bibitem{155} Wolterstorff, \textit{Works and Worlds of Art}, 57-58.
\bibitem{156} van Inwagen, “Existence, Ontological Commitment, and Fictional Entities,” 150 n. 19.
\bibitem{157} Wolterstorff, \textit{Works and Worlds of Art}, 144.
\end{thebibliography}
Goethe’s- Faust kind. Confusingly, Wolterstorff adds that there is a third natural kind, the Faust character simpliciter, that is “included” in the other two natural kinds. Amie Thomasson has pointed out how postulating kinds like Faust simpliciter lead to many problems for establishing identity conditions for fictional entities across texts.

Setting that problem aside, the notion of fictional entities as natural kinds deserves more examination. If fictional entities are seen as natural kinds, the question naturally arises, what are their examples? The examples of a literary work like Dead Souls are all of the individual, physical copies of Dead Souls throughout the libraries of the world, as well as all of the individual readings or utterances of Dead Souls. One may be tempted to think that the examples of the kind Chichikov-in-Dead-Souls, are the “many Chichikovs” that appear individually in the many copies and many utterances of Dead Souls. In other words, one example of the Chichikov natural kind is the Chichikov in Wolterstorff’s copy of Dead Souls, another example of the Chichikov natural kind is the Chichikov in the Rutgers library’s copy of Dead Souls, and so forth. But this is a mistake. The Chichikov that appears in the Rutgers library’s copy of Dead Souls is not an example of a natural kind, rather is the natural kind itself.

A thing would be an example of the Chichikov-in-Dead-Souls if and only if that thing has all of the properties that are ascribed to Chichikov in Dead Souls. The same would apply to any fictional entity. And since it is virtually impossible to find a non-fictional entity that has exactly the same properties as a fictional entity, fictional entities

---

158 Wolterstorff, Works and Worlds of Art, 148.
159 Wolterstorff, Works and Worlds of Art, 148-149.
160 Thomasson, Fiction and Metaphysics, 59-60.
161 Wolterstorff, Works and Worlds of Art, 38.
162 Thomasson, Fiction and Metaphysics, 58.
are, as a rule, unexemplified natural kinds. They are akin to musical works that are never performed and natural kinds like unicorns and hippogriffs.  

Since fictional entities are natural kinds, they exist. Wolterstorff takes as a guiding principle that “one cannot project, single out, refer to, mention what does not exist . . . something cannot have a property at a certain time unless it exists at that time.” Here is where Wolterstorff’s Platonist theory of fictional realism sharply departs from Meinongian and neo-Meinongian fictional realism. Meinongians hold that fictional entities are non-existent, although they are still objects and still have properties. Wolterstorff holds that fictional entities are existent, necessarily and eternally existent in fact, and that it is impossible for a non-existent thing to have properties.

Regarding the nature of fictional entities, Wolterstorff takes it as obvious that they have properties. Fictional entities have all of the properties that are ascribed to them by their stories. Hamlet has the properties of being a prince, being the son of Gertrude, and stabbing Polonius because his story, *Hamlet*, explicitly ascribes these properties to him. Fictional entities also have properties that are merely implied by their stories, but not explicitly stated. Observe the following passage from *Dead Souls*:

> While the server was still spelling out the note, Pavel Ivanovich Chichikov himself set out to look the town over.

This passage concerns, among others, a fictional entity that we could call “Chichikov’s server.” Wolterstorff points out that nowhere in this passage or in *Dead Souls* is it ever

---

166 Wolterstorff, *Works and Worlds of Art*, 146, 149.
stated that Chichikov’s server has the property of being able to see well enough to read and write. Still “we can surely conclude from what Gogol indicates” that Chichikov’s server has this property.

Because Wolterstorff construes fictional entities as natural kinds, an odd consequence follows. All of a fictional entity’s properties will be essential to it. While surely fictional entities have some essential properties (for instance, it seems that being male and being a prince are essential properties of Hamlet), it is counter-intuitive to claim that every property attributed to a character by its story is essential. According to *Hamlet*, Hamlet has the property of attending the university in Wittenberg. Surely if Shakespeare had decided to write that Hamlet attended the university in Cologne, but kept everything else exactly the same in *Hamlet*, then the play would still concern the same character Hamlet. But under Platonist fictional realism, a Hamlet with the property of attending the university in Cologne would not be identical to the Hamlet with the property of attending the university in Wittenberg. Had Shakespeare changed even this most minor of details about *Hamlet*, his play would have featured a completely different natural kind than it, in fact, does.

Ideally a theory of fictional realism should address the fact that fictional characters sometimes appear to violate the law of non-contradiction, and almost always are radically incomplete. Wolterstorff readily concedes that fictional entities are sometimes ascribed contradictory properties by their stories. When this happens, such

---

169 Wolterstorff, *Works and Worlds of Art*, 120.
fictional entities will be impossible kinds. Wolterstorff defines an impossible natural kind as follows:

\[
K \text{ is an impossible kind } \equiv \text{it is impossible that there should be something which is an example of } K. \tag{172}
\]

For instance, the fictional entity of Tully from Everett’s *Asymmetryville* would be an impossible kind.\(^{173}\) According to his story, Tully has the contradictory properties of being identical to Cicero and not being identical to Cicero. Tully is thus an impossible kind because it is impossible for there to be any real thing that is an example of Tully; no real thing can have the property of being identical to Cicero and the property of not being identical to Cicero. But Tully is also a natural kind in virtue of being a fictional entity; all fictional entities are natural kinds. The fact that Tully is an impossible kind does not diminish his status as a natural kind. Wolterstorff reminds us that “the concept . . . of an impossible kind is not that of a kind which cannot exist, but of one which cannot have an example.”\(^{174}\)

Wolterstorff also readily concedes that every fictional entity is radically incomplete, or “non-determinate” with respect to many properties; Chichikov, for instance, is non-determinate with respect to the property of having had a toothache when four years old.\(^{175}\) As a result, Chichikov is a non-determinate natural kind. If a kind “\(K\) is non-determinate, then there is some property \(P\) such that neither having \(P\) nor lacking \(P\)

\(^{172}\) Wolterstorff, *Works and Worlds of Art*, 54.
\(^{173}\) Everett, “Against Fictional Realism,” 634.
\(^{175}\) Wolterstorff, *Works and Worlds of Art*, 147.
is essential within K."176 Just as with impossible kinds, being a non-determinate kind does not diminish a fictional entity’s status as a natural kind. Non-determinate kinds are still natural kinds.

As for the role of the author, it will probably come as no surprise at this point that Wolterstorff’s theory of fictional realism rejects the intuitive picture of an author as the creator, akin to a parent or a god, of her fictional entities. Fictional entities, due to their status as natural kinds, exist necessarily and exist eternally. No author can bring an eternal and necessary entity into existence, nor can any author determine the nature of such an entity. In everyday discourse about fiction, authors are frequently called creative or praised for their creativity. Wolterstorff cautions us against taking such discourse literally.

Insofar as the kinds an author delineates are different from any which he believes to have been exampled or delineated, we may say that he has been creative in his delineation of that character. But to be thus creative is not to bring the character into existence. It is not to create it. Neither is it to bring it about that the character has an example. From the infinitude of person-kinds the author selects one. His creativity lies in the freshness, the imaginativeness, the originality, of his selection, rather than in his bringing into existence what did not before exist.177

Despite the differences in terminology, Platonist fictional realism takes exactly the same position as Meinongianism and neo-Meinongianism regarding the relationship between fictional entities and their authors. Both Platonist fictional realism and Meinongian fictional realism reject folk intuition’s view of the role of the author.

2.5 Problems for Meinongian and Platonist Fictional Realism

Meinongian, neo-Meinongian, and Platonist fictional realism use very different terminology and very different theories of predication from each other. Hopefully these differences do not give the impression that Meinongianism, neo-Meinongianism, and Platonism are radically divergent takes on fictional realism. For in reality, the differences between Meinongian fictional realism and Platonist fictional realism are relatively minor. Ultimately these two theories of fictional realism are quite similar in spirit. Both types of fictional realism hold fictional works and fictional entities to be, for all intents and purposes, the same kind of entities—abstract, timeless objects. But because of this common belief, both types of fictional realism share the same flaws. This section will examine the difficulties raised by Meinongian and Platonist theories of fictional realism. I believe these flaws are serious enough that Meinongian and Platonist fictional realism should be rejected and new theories of fictional realism sought out.

The first common flaw of Meinongian and Platonist theories of fictional realism is their insistence that both fictional works and fictional entities are not physical. Meinong himself appears to take it for granted that literary works and fictional entities are nonphysical. Had Meinong believed fictional works and fictional entities were physical, he would have viewed them as existent, instead of non-existent, objects. Parsons appears to follow suit. He begins his analysis of fictional entities with the unquestioned assumption that they are nonexistent, and thus non-physical. Wolterstorff’s belief that fictional works and fictional entities are natural kinds obviously leads to the conclusion that

---

178 For the sake of simplicity of discourse, I will use the term “Meinongian” to denote both Meinongian and neo-Meinongian views.
fictional works and fictional entities are non-physical. For Platonic entities, like natural kinds, are the paradigmatic example of abstract entities.

And at first it seems quite reasonable to hold that fictional entities and fictional works are not physical objects. To dispute this would be to adopt the infamous “physical object hypothesis” of art.\textsuperscript{179} This hypothesis which, to the best of my knowledge, has never been advocated in its entirety by any philosopher, states that all works of art are physical objects. In the cases of some works of art, this is obviously true. The \textit{Mona Lisa} is a physical object; it is a certain piece of canvas, covered with paint, and hanging in the Louvre.

But, as Richard Wollheim points out in \textit{Art and its Objects}, the physical object hypothesis fails miserably when it tries to account for art works like musical compositions and literature.\textsuperscript{180} It is very difficult to imagine what physical object is to be identified with a specific literary work such as \textit{Ulysses}. Wollheim says that if we identify \textit{Ulysses simpliciter} with the physical object that is Wollheim’s personal copy of \textit{Ulysses}, absurd consequences follow. If Wollheim’s copy of \textit{Ulysses} were to be lost or destroyed, then \textit{Ulysses} itself would be lost or destroyed.\textsuperscript{181} This will be the case even if we identify \textit{Ulysses} itself with James Joyce’s original manuscript.\textsuperscript{182} Nor is it wise to identify \textit{Ulysses} the work with the class of all of its physical copies. The class of all physical copies of \textit{Ulysses} is constantly expanding, as the novel remains in publication. This class has expanded considerably since Joyce’s death. If \textit{Ulysses} is to be identified with this

\textsuperscript{179} Wollheim, \textit{Art and its Objects}, 4.
\textsuperscript{180} Wollheim, \textit{Art and its Objects}, 4-10.
\textsuperscript{181} Wollheim, \textit{Art and its Objects}, 5-7.
\textsuperscript{182} Wollheim, \textit{Art and its Objects}, 7.
class, then we cannot say that Joyce completed *Ulysses* in his lifetime.\(^{183}\) This is just one of several counter-intuitive consequences of identifying a literary work with the class of all its physical copies.\(^{184}\)

There are other reasons to scoff at the notion that fictional works should be identified with physical objects. Some fictional works, folk tales in particular, are never written down, or published, or recorded in some other physical way. There is no physical object that can plausibly be identified with an oral folk tale. Nonetheless, these kinds of folk tales seem to qualify as fictional works. Even when a folk tale is eventually put into writing, and a physical copy of it comes into existence, this does not change the folk tale into a fictional work for the first time. Rather the written form of a folk tale is properly viewed as a copy of or version of the folk tale, which was already a fictional work in its non-physical, oral form.

A fictional work can also exist solely within a person’s memory. It is conceivable that a person with an exceptionally prodigious memory could memorize a fictional work like *Pride and Prejudice* word-for-word. Within such a person’s mind would be a “copy” of sorts of *Pride and Prejudice*, but it would not be what most people would call a physical copy.\(^{185}\) Now if by some unusual circumstances, all physical copies of *Pride and Prejudice* were destroyed, and the only copy of the work left was the memorized, non-physical copy, it seems *Pride and Prejudice* would still exist. This is the kind of scenario found in Ray Bradbury’s *Fahrenheit 451*. In *Fahrenheit 451*, all of the world’s

---
\(^{184}\) Wollheim, *Art and its Objects*, 8-10.
\(^{185}\) At least at the level of folk intuition few people would call a memorized copy a physical copy. Debates within philosophy of mind over whether mental phenomena are best viewed as physical phenomena do not, in my experience, influence the folk to a great extent. A sort of casual Cartesian dualism seems to be quite widespread among non-philosophers.
books have been burned, and no physical copies of fictional works remain. But a small
group of persons, blessed with prodigious memorization powers, have memorized the
texts of many fictional works; this keeps the works in existence. The world of
*Fahrenheit 451* is a world without books, but it is not a world without fiction.

If it is ridiculous to identify fictional works with physical objects, it seems that it
is also equally ridiculous to identify fictional entities with physical objects, perhaps even
more so. It is somewhat understandable how one could fall into the mistake of thinking a
work like *Ulysses* is identical to the physical copy on one’s coffee table. But it is hard to
see how one could mistakenly identify a fictional entity like Sherlock Holmes with a
physical object. It is difficult to even imagine what physical objects potentially would be
candidates for this misidentification. Maybe Sherlock Holmes could be identified with a
real man who worked as a detective in London at the turn of the last century? No, even
this is a stretch. Folk level intuitions about ontology are sometimes quite mistaken, but
rarely are they this skewed.

I think it is safe to conclude that fictional works and fictional entities are not
identical to any sort of physical object. Still I believe that Meinongian, neo-Meinongian,
and Platonist theories of fictional realism err greatly when they hold fictional works and
entities to be purely abstract entities. While fictional works and entities are not physical
objects themselves, they stand in *extremely* important relations to physical objects.
Specifically both fictional works and fictional entities seem to depend on physical objects
in order to remain in existence.

In his argument against the physical object hypothesis, Wollheim points out that a
physical copy of *Ulysses* can be destroyed without the work *Ulysses* being destroyed.
This fact is seen as intuitively true, and Wollheim uses it to show that the physical object hypothesis is counter-intuitive. However I am not sure if our intuitions about the non-physicality of fictional works and fictional entities can be sustained if we elaborate further on Wollheim’s example. It seems clear that if one physical copy of *Ulysses* is destroyed, *Ulysses* itself survives. It also seems clear that even if Joyce’s original manuscript is destroyed, *Ulysses* the work would still exist. But now imagine that due to some *Fahrenheit 451*-like scenario, every physical copy of *Ulysses* in the entire world has been irretrievably destroyed. Every paper printed copy is gone, every audio copy is gone, and every electronic copy is gone. In this case, would we still believe that *Ulysses simpliciter* has survived? While it is always difficult to say definitively what folk intuition would be in such a thought experiment, I am inclined to think that most people would say that in this scenario, *Ulysses simpliciter* has *not* survived. In such a world, *Ulysses* does not exist anymore. *Ulysses* may not be identical to any physical object or class of physical objects, but intuitively it seems that the existence of *Ulysses* depends on the existence of some physical object or objects.

Drawing on the *Fahrenheit 451* scenario, imagine that despite the destruction of all physical copies of *Ulysses*, one person, let’s call him Granger, has memorized *Ulysses* verbatim. Intuitively it seems that *Ulysses* still exists because Granger’s memorized “copy” exists. The fact that a memorized copy of a novel hardly qualifies as a physical object is supposed to show that the physical object hypothesis fails. But upon closer examination, this case actually indicates that fictional works bear an important relationship to some physical objects. Setting aside the difficult question of whether or not the contents of Granger’s mind (such as the memorized copy of *Ulysses*) are properly
seen as physical, Granger himself surely is a physical object. If in the *Fahrenheit 451* scenario, *Ulysses* exists because a memorized copy of it exists in Granger’s mind, then it will also be the case that *Ulysses* exists because Granger, a physical object, exists. And if Granger himself were to be killed, taking the only copy of *Ulysses* to his grave, so to speak, I think our intuitions would say that *Ulysses* no longer exists. *Ulysses* may not be a physical object itself, but it seems to require the existence of some physical objects, whether printed copies or human beings, in order to stay in existence.

The same principles apply to the case of folk tales. Folk tales are generally memorized by several persons in a cultural group and passed on to other members of the group through oral recitation. As long as those people who have heard the folk tale memorize it and recite it to others who have not yet heard it, and those others in turn memorize the tale and recite it anew, the folk tale can remain in existence for centuries after the deaths of the original tellers. People, specifically people who engage in specific sorts of physical actions involving physical organs like ears and vocal chords, actions like memorizing and reciting, keep a folk tale in existence. And if, for some reason, all of the persons who have a given folk tale memorized were to die out without reciting and passing on the tale to a new generation, it seems that the folk tale would no longer exist.\(^{186}\) In virtue of depending on people in order to exist, folk tales depend on physical objects to exist.

Fictional works apparently depend on certain physical objects in order to exist. Consequently, fictional entities, which intuitively are seen as “in” fictional works or in some way dependent upon fictional works, will also depend on certain physical objects in order to exist. If the existence of *Ulysses* depends upon the presence of physical copies

\(^{186}\) Thomasson notes this in *Fiction and Metaphysics*, 9.
or a physical person who has memorized the text, it would seem that fictional entities from *Ulysses* like Molly Bloom also depend on the presence of physical copies or a physical person with a good memory in order to exist. If every copy of *Ulysses* were to be destroyed and Granger to be killed, it is difficult to imagine how Molly Bloom would still exist.

The dependence that fictional works and fictional entities have on physical objects is a serious problem for Meinongian, neo-Meinongian, and Platonist theories of fictional realism. Meinongianism and neo-Meinongianism construe fictional works and entities as non-existent objects. Meinongian non-existent objects are eternal, necessary objects. Such objects cannot depend on some other entity for their object-hood, let alone on physical, existent entities like people and printed books. Platonist fictional realism construes fictional works and entities in the same manner, only asserts that such things exist. A natural kind is an eternal and necessary thing; it too cannot depend on physical entities in order to exist. Yet our intuitions indicate that fictional works and fictional entities, while not being physical objects themselves, do depend on certain kinds of physical objects in order to exist. Meinongianism and Platonism’s insistence upon the non-physicality of fictional works and objects conflicts sharply with ordinary intuitions.

A second flaw shared by Meinongian and Platonist theories of fictional realism is their insistence that fictional works and fictional entities are timeless objects. Meinongian non-existents are both non-spatial and non-temporal. Fictional works and fictional entities, being non-existent objects, are non-temporal. Platonic entities like natural kinds are also timeless objects; they exist from eternity, never coming into
existence at a specific point in time or ceasing to exist. If fictional works and fictional entities are natural kinds, then fictional works and fictional entities exist from eternity.

The notion that fictional works are timeless and do not come into existence at a certain point in time is in serious conflict with ordinary beliefs about fiction and with professional literary scholarship. Ordinary belief holds that *Pride and Prejudice*, for instance, was written in the early 1800s. To say that *Pride and Prejudice* existed before that time period, such as in the fourteenth century, is to say something very strange as far as most readers are concerned. Meanwhile at the level of professional literary scholarship, the proper dating of texts is of utmost importance. One of the largest debates in medieval studies, for example, is whether *Beowulf* was composed in the seventh, eighth, ninth, or tenth century. All scholars party to this debate recognize that these four positions are mutually exclusive; if *Beowulf* indeed was composed in the tenth century then it makes no sense to speak of it existing in the seventh century.

But according to Platonism and Meinongianism, ordinary intuition and professional scholarship are wrong about these matters. *Pride and Prejudice* indeed existed in the fourteenth century (or was an object in the fourteenth century, to put it in Meinongian terms). *Pride and Prejudice* has always existed; even at the moment of the Big Bang, the abstract furniture of the universe, so to speak, included *Pride and Prejudice*. *Beowulf* too existed at the moment of the Big Bang (or was an object at the moment of the Big Bang). Contrary to what scholars of medieval literature might believe, even if *Beowulf* was composed in the tenth century, it still existed in the seventh century.

---

century. Whatever activity may have happened in the tenth century was merely the
discovery of an entity that had been in existence (or had been an object) from eternity.

Meinongianism and Platonism’s view of literary works as timeless runs into
additional problems. Ordinary intuition and professional literary scholarship commonly
use terms like avant-garde, reactionary, typical of the era, and the like to describe
fictional works. These terms seem to be quite meaningful; they indicate important,
perhaps essential, properties of fictional works. But terms like reactionary and avant-
garde also presuppose that fictional works come into existence at a particular point in
time and stand in relationships to other works which also come into existence at
particular points in time. *Ulysses* is considered avant-garde because its use of language is
not found in any other novels contemporary with or prior to *Ulysses*, but such use of
language is found among novels written after *Ulysses*. But under Meinongianism or
Platonism, a term like avant-garde is meaningless. Literary works are timeless entities.
There are no novels that came “before” *Ulysses* from which *Ulysses* could depart
radically in use of language, and there are no novels that came “after” *Ulysses* which
*Ulysses* could influence. *Ulysses*, and all other novels have existed (or have been objects)
for eternity.

Satirical literature presents an especially difficult case for Meinongian, neo-
Meinongian, and Platonist fictional realism. Satirical fictional works often parody real
life contingent objects like people, events, and institutions. *Animal Farm* is well-known
as a satire of the Soviet Union, and being a satire of the Soviet Union is probably an
essential property of *Animal Farm*.\(^{189}\) The Soviet Union was a contingent entity that

\(^{188}\) Thomasson, *Fiction and Metaphysics*, 9.

\(^{189}\) Thomasson *Fiction and Metaphysics*, 8-9.
came into existence at a specific point in time and went out of existence at a specific point in time. Yet if Meinongianism or Platonism is correct, *Animal Farm* is a necessary object that has existed (or been an object) from eternity. Since eternal objects are immutable, *Animal Farm* has always had the property of being a satire of the Soviet Union. But it seems quite strange indeed to think that in the year 917 A.D., a millennium before the Soviet Union came into existence, there already existed (or already was an object) a fictional work that had the property of being a satire of the Soviet Union. Granted, this scenario is not *prima facie* impossible, but it is very odd and counter-intuitive.

The insistence of Meinongian and Platonist fictional realism that fictional works are outside of time clashes with ordinary intuitions about fiction as well as beliefs that are foundational for literary scholarship. Due to the strong relationship between fictional works and fictional entities, the counter-intuitive claims of Meinongian and Platonist fictional realism will also apply to fictional entities. Fictional entities exist (or are objects) from eternity. Contrary to ordinary and scholarly belief, Hamlet has always existed (or been an object). Hamlet existed in 1464, a century before Shakespeare himself existed. *Animal Farm*’s Snowball, who seems to have an essential property of being a parody of Trotsky, had the property of being a parody of Trotsky countless centuries before Trotsky himself came into existence.

The final, and in my view, most serious shared flaw of Meinongian, neo-Meinongian, and Platonist forms of fictional realism is their view on the role of the author. Ordinary intuitions about fiction as well as literary scholarship construe the
author as a very important figure with regards to fictional works and fictional entities. In ordinary, pre-theoretic discourse about fiction it is very common to speak of authors as creators of fictional characters and the fictional worlds the characters inhabit. It is not unusual to hear authors compared to parents or even gods. Intuitively it seems that authors cause fictional entities to come into existence and cause those fictional entities to be the way that they are.¹⁹⁰

I imagine that most ordinary readers would agree that if an author had not come into existence himself, the fictional entities in his works would never have existed either. If Shakespeare had never been born, then Hamlet, Ophelia, and Polonius would never have existed. I also imagine that most ordinary readers would agree that if an author had never written the fictional works that he wrote, then the fictional entities in those works would never have existed. If Arthur Conan Doyle had been much busier with his medical practice and had no spare time to write any Sherlock Holmes stories, then Sherlock Holmes would never have existed. And I imagine that most ordinary readers would agree that if an author had decided to write her fictional works differently then the fictional entities in those works would be different than they, in fact, are. If Austen had written Pride and Prejudice but decided to make the character Lydia Bennett shy and prudent instead of flirtatious and feckless, then Lydia Bennett would, in fact, be shy and prudent.

We also see the intuitive connection between the author and the nature of fictional entities in how radical interpretive statements about fiction are received at the folk level. From my own experiences in high school and college English classes, I observed that the

¹⁹⁰ Interestingly, folk-intuitions about the created status of literature may not apply to other works of art. In “Platonism in Music: A Kind of Defense,” Grazer Philosophische Studien 19 (1983), Kivy notes that now common the view of the artist as a godlike creator is a relatively recent development, especially in the case of musical works. “Bach did not think he was God; Wagner did, with dire results” (115).
overwhelming folk response to “radical” interpretive statements about fictional entities, such as “Hamlet has an Oedipus complex,” is skepticism. Those who consider it ludicrous to claim that Hamlet has an Oedipus complex invariably appeal to Shakespeare himself. They will argue that Hamlet could not possibly have an Oedipus complex because an Oedipus complex, as everyone knows, is a Freudian concept, and Shakespeare died centuries before Freud was born. Even if Shakespeare had wanted to give Hamlet the property of having an Oedipus complex, he would not have been able to. Since Shakespeare did not give Hamlet the property of having an Oedipus complex, Hamlet does not have this property - for where else could Hamlet get his properties except from his author Shakespeare?

Regardless of how well this line of reasoning would be received by scholars of literary theory, it is evidence for an important folk intuition about the relationship between an author and the properties of fictional entities. According to ordinary intuition, fictional entities have specific properties and lack specific properties. The author ultimately determines which properties a fictional entity has and lacks. If for some reason an author could not have or would not have given a certain property to a fictional entity, then that fictional entity cannot have that property.

A theory of fictional realism should have a robust and commonsensical position on the role of the author. Meinongianism and Platonism do not have this. The position these theories of fictional realism take is intolerably counter-intuitive. Combined with the counter-intuitiveness of Meinongianism and Platonism’s claims that fictional works and fictional entities are non-spatial and non-temporal, there are just too many counter-intuitive consequences of these views. For this reason I cannot be a Meinongian, neo-
Meinongian, or Platonist fictional realist. Let us now look elsewhere to find a theory of fictional realism that is more palatable to our intuitions about the spatio-temporal dependencies of fictional entities and our intuitions about the role of the author in creating fictional works and entities and determining some of their properties.

2.6 Croce and Collingwood

In our quest for a theory of fictional realism that has a robust conception of the author, we would do well to consider what two early twentieth century philosophers, Benedetto Croce and R.G. Collingwood, have to say about fictional works and fictional entities. Neither of these philosophers can properly be classified as fictional realists. They take no explicit positions on whether or not fictional entities are existent; rather they freely speak of fictional works and fictional entities as existing, apparently unaware of how philosophically loaded the term “exists” in light of Meinong’s work. And while Croce and Collingwood do offer positions on what kind of objects fictional entities are, they do not develop ontologies for fictional entities to the extent that metaphysicians like Parsons, Wolterstorff, and Thomasson do.

Nonetheless it is important to study Croce and Collingwood on fiction. Their position on the role of the author is, in a sense, the polar opposite of the position that Meinongianism and Platonism take on the role of the author. Since Meinongianism and Platonism’s views on the author are, in my opinion, grounds for rejecting these forms of fictional realism, it will be quite instructive to see how theories of fiction with the opposite views hold up in the search for a plausible theory of fictional realism.
Had this dissertation been written before World War II, the figure of Benedetto Croce would hardly require introduction. In the early twentieth century, Croce (1860-1952) was one of Europe’s most famous philosophers. His works on aesthetics, history, and literary criticism were widely read.\footnote{Gary Kemp, “Croce’s Aesthetics,” The Stanford Encyclopedia of Philosophy (2008).} Croce was also active in Italian politics; shortly after the war he was seriously encouraged to seek the presidency of Italy.\footnote{Cecil Sprigge, introduction to Philosophy, Poetry, History: An Anthology of Essays, by Benedetto Croce (London: Oxford University Press, 1966) xxviii.} But Croce’s ideas and writing style did not age well as trends changed in both continental and analytic philosophy, leaving him largely forgotten among philosophers today.\footnote{Kemp, “Croce’s Aesthetics,” 1.}

In 1902, two years before Meinong’s “The Theory of Objects” was published, Croce put forth his own views about the nature of art works in *The Aesthetic as the Science of Expression and of the Linguistic in General*. As its lengthy title suggests, this was a very comprehensive work. It is not merely a treatise on aesthetics or on philosophy of language, but rather an explication of Croce’s systematic philosophy. In this work, Croce adopts a form of idealism and argues that the only thing that exists is “lived human experience, taking place concretely at particular times and places.”\footnote{Kemp, “Croce’s Aesthetics,” 3.} Since human experience alone exists, it is obviously quite important to study the nature of human experience, which is what Croce does in *The Aesthetic*. He begins by noting that human beings are constantly presented with an endless flux of sensory perceptions. We perceive sounds, tactile feelings, tastes, colors, shapes, and objects; Croce calls these perceptions of various stimuli “intuitions.”\footnote{Benedetto Croce, *The Aesthetic as the Science of Expression and of the Linguistic in General*, trans. Colin Lyas (New York: Cambridge University Press, 1992), 7.} It might seem that perception is a passive process;
humans receive intuitions much like a stone soaks up the rays of the sun. But Croce argues that human perception is an active process. The “human spirit,” or the mind, receives intuitions and actively imposes form and order on them in order to produce an idea or notion of the world. Croce calls the mental image that the mind forms in response to sensory stimulation a “representation” or alternately an “expression.”

This might sound like a description of what an artist does. For instance, a sculptor is presented with a block of marble. Using his creative powers, the sculptor actively imposes a form on the marble, resulting in a statue of Bacchus. The statue is a representation of something, Bacchus, but of course it is not identical to Bacchus himself. We can also speak of how the statue “expresses” the artist’s idea of what Bacchus looks like, and we can note how each artist’s idea of Bacchus is unique. Michelangelo’s Bacchus is quite different from Sansovino’s Bacchus because Michelangelo’s mental picture of Bacchus is unique and different from Sansovino’s unique mental picture of Bacchus.

The similarity between an artist’s work and Croce’s account of human perception is no coincidence. According to Croce, all acts of ordinary human perception are acts of artistic creation and representation, albeit on a mundane level. Just as the sculptor is presented with a block of marble, raw material, we are presented with intuitions. Just as the sculptor actively imposes a form on the marble resulting in Bacchus, we actively impose form and order on our intuitions resulting in a mental image of the world. Just as the statue is a representation of Bacchus (but not Bacchus himself), our mental images of

196 Colin Lyas, translator’s forward to The Aesthetic as the Science of Expression and of the Linguistic in General, by Benedetto Croce (New York: Cambridge University Press, 1993), xix.
197 Croce, The Aesthetic as the Science, 6-7; Lyas, translator’s forward, xx.
198 Croce, The Aesthetic as the Science, 8-9, 11; Lyas, translator’s forward, xx.
the world are representations of the world (but not identical to the world itself). And just as the statue expresses the artist’s unique idea of what Bacchus looks like, our mental image of the world expresses our unique notion of what the world is like. To perceive the world is to be an artist, Croce says. “Each of us, in short, is in some small measure a painter, sculptor, musician, poet, prose writer.”

This is a startling claim indeed. Ordinarily it is thought that most people are not artists; persons with true artistic talent are quite rare. Great composers like Bach, great sculptors like Michelangelo, and great poets like Dante, are revered because it is thought that they engage in a rarified kind of activity, artistic creation, that very few ordinary people are capable of. But according to Croce, the activities of great artists differ from those of ordinary people by degree, not in kind. The people that we praise as “artists” are praised because they have a stronger sense of perception, a better eye for detail, so to speak, than the average person does.

Artists are also distinguished from the general population by their inventive responses to sensory stimuli. “The world that we normally intuit is a petty thing and translates itself into petty expressions.” Few people, when presented with the range of sounds audible to the human ear, can arrange these into a new piece of music that hitherto did not exist, let alone a symphony. Few people, when presented with visual stimuli of colors, shapes, and textures, can arrange these into a painting, let alone a masterpiece. “But, from time to time . . . we pass . . . from petty intuitions to the greater, and so to the

---


sublimest and greatest.“ This is what great artists like Bach, Raphael, and Dante have done. And since “the passage is sometimes far from easy” we rightly praise great artists as rare and wonderful. 

All mental representations or expressions are artistic, according to Croce. The people we praise as great artists are people who create very beautiful, detailed, and inventive mental representations. And the objects that we praise as great artworks - Michelangelo’s Bacchus, Beethoven’s Ninth Symphony, Dante’s Divine Comedy, etc - are very beautiful, detailed, and inventive representations. This last claim is central to Croce’s philosophy. According to Croce, every work of art is a mental representation or expression created within the mind of the artist. Pride and Prejudice is a mental expression within the mind of Jane Austen; it is not the physical copy sitting on my bookshelf. The Ninth Symphony is a mental expression within the mind of Beethoven; it is not a printed score or a performance by any orchestra. And surprisingly, Michelangelo’s Bacchus is a mental expression within the mind of Michelangelo; it is not the sculpted mass of marble standing in the Bargello. Croce thoroughly rejects the physical object hypothesis. Art works are identical to a mental representation within the mind of an artist. They are never identical to a physical object out in the world. “The work of art (the aesthetic work) is always internal; and what is called external is no longer a work of art.”

What then are we to make of objects like the sculpted mass of marble in the Bargello or the physical copy of Pride and Prejudice on my shelf? Croce says that such

---

202 Croce, The Aesthetic as the Science, 10.
203 Croce, The Aesthetic as the Science, 10.
objects are not art works proper. Rather artists create objects like sculptures and paintings, and have musical scores and literary works published in physical form in order to communicate their mental representations, which are art works, to other people.

And what else are those combinations of words called poetry, prose, poems, novels, romances, tragedies or comedies, but physical stimulants of reproduction? The spiritual energy of memory, with the assistance of the physical facts above mentioned, makes possible the preservation and production of the intuitions produced by man.\(^\text{205}\)

And while artists generally produce a physical stimulant of reproduction, like a manuscript or painting, this is not a necessary procedure for creating an art work. Once created, an art work exists entirely and fully formed within the mind of an artist.

The aesthetic fact is altogether completed in the expressive elaboration of impressions. When we have achieved the word within us, conceived definitely and vividly a figure or a statue, or found a music motive, expression is born and is complete; there is no need for anything else.\(^\text{206}\)

If an artist chooses to communicate his mental representation with other people by producing a physical object of some sort, this is merely icing on the cake, so to speak.

The poem, we must see, is there already in its entirety when the poet has expressed it in words uttered to himself. When he goes on to raise his voice to utter it to others, or procures that others shall commit it to memory and chant it to each other as in a schola cantorum, or shall record it in writing and printing, this is a new stage of great social and cultural importance having not an aesthetic but a practical character.\(^\text{207}\)

\(^{205}\) Croce, *The Aesthetic as the Science*, 108.

\(^{206}\) Croce, *The Aesthetic as the Science*, 56.

Physical stimulants of reproduction are certainly valuable, for they allow other people to enjoy what would remain the private mental representation of an artist. But these physical stimulants are unnecessary to the process of creating an art work itself. In fact, Croce points out, an artist would not be able to create a physical stimulant of reproduction unless the art work already exists in his mind.

The painter, who paints upon canvas or upon wood, but could not paint at all, did not the intuited image, the line and colour as they have take shape in the fancy, precede, at every stage of the work, from the first stroke of the brush or sketch of the outline to the finishing touches, these manual actions.\(^{208}\)

Croce’s philosophy encompasses all forms of art, including fictional works. Fictional works like *Pride and Prejudice*, *Hamlet*, and *Ulysses* are mental representations within the minds of their respective authors. Presumably, fictional entities like Lydia Bennett, Hamlet, and Molly Bloom are also mental representations within the minds of the authors, although Croce never discusses the status of fictional entities specifically.

It is not difficult to see how sharply Croce’s account of the author contrasts with Meinongian, neo-Meinongian, and Platonist fictional realism. According to Meinongianism and Platonism, authors do not create fictional works and fictional entities, but rather discover them or select them. But according to Croce, authors actively create fictional works and fictional entities. Meinongianism and Platonism also claim that fictional works and entities are necessary objects that exist (or are objects) either among the infinite collection of non-existents or in Plato’s heaven. Croce rejects this view completely. Fictional works and entities are contingent objects, since they are brought into existence only by the deliberate activities of human authors. And fictional

\(^{208}\) Croce, *Philosophy, Poetry, History*, 228; see also *The Aesthetic as the Science*, 56, 114-115.
works and entities do not dwell in some distant realm of eternal abstracta. Instead works of art “exist nowhere else than in the souls which create or recreate them.”

Let us now turn our attention to the British philosopher R.G. Collingwood (1889-1943). Collingwood’s philosophical interests were broad, but he is best remembered today for his work in aesthetics, particularly his 1937 *The Principles of Art*. Collingwood, a follower and early translator of Croce, adopts a position on the nature of art works very similar to Croce’s. But Collingwood reaches this conclusion via first principles that are quite different from Croce’s views.

Croce, as we will recall, argued that all mental experiences involving sensory stimulation are essentially artistic. The difference between the mental expression I form of my neighbor’s trash can and the mental expression that Michelangelo formed and communicated by painting *The Last Judgment* is a difference of degree not kind. Michelangelo’s mental representation is clearly more detailed, beautiful, and inventive than my mental representation of the trash can, but we both engage in the same kind of artistic activity.

Collingwood initially rejects the claim that all human activities should be classified as artistic. His concern in *The Principles of Art* is to answer the simple question “what is art?” To answer this question, we must first consider which sorts of human activities are *not* art, even if they are occasionally referred to as art in ordinary speech. The first such activity, according to Collingwood, is crafts. It is sometimes

---

210 Kemp, “Croce’s Aesthetics,” 1.
thought that art is a craft similar to woodworking, blacksmithing, weaving, or horse-training. Just as a carpenter is a craftsman who produces tables, and a cobbler is a craftsman who produces shoes, the artist is a craftsman who produces paintings, symphonies, or novels. This is completely erroneous, according to Collingwood.

One distinguishing feature of all crafts is “a distinction between means and end, each clearly conceived as something distinct from the other but related to it.” A craftsman makes use of various materials and engages in various activities all as a means to an end. A blacksmith, for instance, acquires iron, fires up his forge, and manipulates his tools in order to produce a horseshoe. The blacksmith has the end, a horseshoe, in mind before he engages in any of the means of production; the preconceived end guides his productive activities. But art, Collingwood argues, *never* involves a distinction between means and an end. When producing a work of art, an artist does not begin with a preconceived end in mind which guides his productive activities. Therefore art is not a craft.

Religious and social rituals, or “magic” as Collingwood calls them, are also not to be classified as works of art, despite how aesthetically attractive such rituals may be. Rituals, from Aboriginal initiation rites to a high-society debutante ball, all aim at the same end - the arousal of specific emotions in the participants and spectators. But this, Collingwood points out, is a distinction between means and an end. Art, lacking the means and end distinction entirely, cannot be magic.

---

214 Collingwood, *The Principles of Art*, 20-21. There are other differences between crafts and art that Collingwood discusses (see 5-17, 21-26) but the distinction between means and an end is the most relevant to our discussion.
Similarly, instances of what Collingwood calls “amusement” – pornography, gory horror novels, dime novels, detective stories, and sentimental “historical” literature – are also not art, despite any aesthetic merits they might have.\textsuperscript{216} These amusements, like magic, are created for the purpose of generating certain emotional states in their audiences, although the emotions generated by amusement are quite different from those generated by magic. Nonetheless, this is a distinction between means and an end, and since art cannot possess this distinction, art cannot be amusement.\textsuperscript{217}

While art cannot have as its purpose arousing certain emotions, there is a deep connection between art and emotions. Collingwood argues that all art works express emotions.\textsuperscript{218} A poem, a novel, a symphony, a statue, a painting, or a dance is an expression of an emotion of its author, albeit a very complex emotion. Now it may seem that a distinction between means and an end is involved here. If \textit{Las Meninas} is an expression of a complex emotion that Velazquez had, then it seems that Velazquez engaged in certain activities (like painting a canvas) in order to produce a specific end (expressing his complex emotion). But this is not what is involved in the expression of emotions. This scenario assumes that before Velazquez put his brush to the canvas, he was already fully aware of the nature of the complex emotion that is expressed by \textit{Las Meninas}. However, according to Collingwood, no one is ever aware of the nature of his emotion before he expresses it.

When a man is said to express emotion, what is being said about him comes to this. At first he is conscious of having an emotion, but not conscious of what this emotion is. All he is conscious of is a perturbation or excitement, which he feels

\begin{thebibliography}{9}
\bibitem{216} Collingwood, \textit{The Principles of Art}, 84-88.
\bibitem{217} Collingwood, \textit{The Principles of Art}, 78-104.
\bibitem{218} Collingwood, \textit{The Principles of Art}, 109.
\end{thebibliography}
going on within him, but of whose nature he is ignorant. While in this state, all he can say about his emotion is: ‘I feel . . . I don’t know what I feel.’ From this helpless and oppressed condition he extricates himself by doing something which we call expressing himself.219

As the artist expresses his emotion, he simultaneously discovers his emotion. All of the characteristics of the emotion, its complexity, its beauty, its inventiveness, are brought to the attention of the artist for the very first time through the process of expression. The artist is simultaneously creator of and audience of the art work.220 “Until a man has expressed his emotion, he does not yet know what emotion it is. The act of expressing it is therefore an exploration of his own emotions. He is trying to find out what these emotions are.”221 The expression of emotion, so conceived, does not involve a distinction between means and an end. The artist cannot begin the expressive process with a preconceived end in mind, such as “express emotion e”, because the purpose of expressing an emotion is to discover e.222 For this reason, art is the expression of emotion rather than ritual or amusement.

Here we see a striking similarity between Collingwood’s views on art and Croce’s. Croce believed that every work of art was an expression of an artist’s notion of the world. Collingwood believes that every work of art is an expression of an artist’s emotion. The expressing of emotion is a mental process which occurs entirely within the mind of an artist. It follows then that the result of this process, an art work, is a mental

221 Collingwood, The Principles of Art, 111.
222 Collingwood, The Principles of Art, 111.
object, or an “imaginary” object to use Collingwood’s terminology.\textsuperscript{223} Regarding music, Collingwood writes

If the making of a tune is an instance of imaginative creation, a tune is an imaginary thing. And the same applies to a poem, or a painting or any other work of art. This seems paradoxical; we are apt to think that a tune is not an imaginary thing, but a real thing, a real collection of noises; that a painting is a real piece of canvas covered with real colours; and so on.\textsuperscript{224}

Despite the pull of these intuitions, Collingwood cautions against identifying a work of art with any real object or real event. The real noises that are produced when a piece of music are performed are “useful accessories” of the art work.\textsuperscript{225} A public performance of a musical work, or even a printed score of the musical work, enables other people to access the emotion expressed by the composer, an emotion that previously existed only within the composer’s mind.

The artist, when he has made his tune, may go on to do something else which at first sight seems to resemble this: he may do what is called publishing it. He may sing or play it aloud, or write it down, and thus make it possible for others to get into their heads the same thing which he has in his. But what is written or printed on music-paper is not the tune. It is only something which when studied intelligently will enable others (or himself, when he has forgotten it) to construct the tune for themselves in their own heads.\textsuperscript{226}

Printed copies of fictional works serve a similar purpose.

\textsuperscript{224} Collingwood, The Principles of Art, 139.
\textsuperscript{225} Collingwood, Outlines of a Philosophy of Art, 12; The Principles of Art, 134.
\textsuperscript{226} Collingwood, The Principles of Art, 134-135.
The object, in the case of art, is an imaginary object, not a real object. Shakespeare’s printed text is a real object, and really lies before me; but to contemplate the tragedy of Hamlet is not to perceive this printed book but to ‘see’ Hamlet himself as Shakespeare ‘saw’ him. This ‘seeing’ is the contemplation of a human character, human words, human actions; but the character, words, and actions of an imaginary human being.\textsuperscript{227}

It is interesting to note that in this passage, Collingwood not only classifies a fictional work, Hamlet, as an imaginary object, but also classifies the fictional entity Hamlet as an imaginary object.

And while the production of printed texts, musical scores, painted canvases, and the like serves a valuable social purpose, such production is not necessary in order to create a work of art. Following Croce, Collingwood notes that once an artist has successfully expressed his emotion and created a work of art, this work of art exists fully and completely within the artist’s mind. The work of art exists regardless of whether or not it can be shared with other people via some physical object or performance. Speaking of an artist who creates a piece of music, Collingwood writes

This tune is already complete and perfect when it exists merely as a tune in his head, that is, an imaginary tune. Next, he may arrange for the tune to be played before an audience. Now there comes into existence a real tune, a collection of noises. But which of these two things is the work of art? Which of them is the music? The answer is implied in what we have already said: the music, the work of art, is not the collection of noises, it is the tune in the composer’s head.\textsuperscript{228}

Like Croce, Collingwood’s views on fiction are in direct opposition to Meinongianism and Platonism. Fictional works and fictional entities are not eternal, necessary entities. Rather they are contingent entities, brought into existence at a certain point in time, only

\textsuperscript{227} Collingwood, \textit{Outlines of a Philosophy of Art}, 12.
\textsuperscript{228} Collingwood, \textit{The Principles of Art}, 139.
through the deliberate creative activity of their authors. Collingwood is particularly insistent upon the notion of an artist as creator of art works. “In order that a work of art should be created, the prospective artist . . must have in him certain unexpressed emotions, and must also have the wherewithal to express them.”

He considers the relationship between a work of art and its author as akin to the relationship between a child and its parent.

To create something means to make it non-technically, but yet consciously and voluntarily. Originally creare means to generate, or make offspring, for which we still use its compound ‘procreate,’ and the Spaniards have criatura, for a child. The act of procreation is a voluntary act, and those who do it are responsible for what they are doing . . . It need not be done (as it may be in the case of a royal marriage) as a means to any preconceived end. It need not be done . . . according to any preconceived plan. . . . It should be clear that when we speak of an artist as making a poem, or a play, or a painting, or a piece of music, the kind of making to which we refer is the kind we call creating.

Both Croce and Collingwood offer accounts of the author that are much more amenable to ordinary intuition than Meinongianism and Platonism’s account of the author. Idealist fictional realism rescues fictional entities and fictional works from the esoteric realms of Meinongian non-existents and Plato’s heaven and places them securely within the minds of their authors. Fictional entities and works are no longer necessary and eternal entities, waiting to be discovered like some far off galaxy. They are contingent entities that exist only because their authors have decided to create them.

229 Collingwood, The Principles of Art, 130.
230 Collingwood, The Principles of Art, 128-129.
2.7 Problems for Idealist Fictional Realism

Idealist fictional realism enjoys a decisive advantage over Meinongianism and Platonism in virtue of the robust role it attributes to the author in the creation of fictional entities and works. However, binding fictional entities and works tightly to the mental activities of their authors leads to problems of its own.

Both Croce and Collingwood assert that works of art, including fictional works and fictional entities, exist completely and fully formed within the mind of their artists. While artists often create concrete objects like paintings or printed books, this is only done to communicate their art works to other people. These physical objects are not identical to any work of art, and it is not necessary for an artist to create a physical object in order to create a work of art. A work of art already exists, with all its details, within the mind of an artist.

This will strike many as a deeply implausible view of art, particularly with regards to art works like paintings and sculptures. Ordinarily we think of a painting or a sculpture as a physical object composed out of a specific physical medium; the notion of a painting that exists only in the mind of an artist will strike many people as outright incoherent. It is especially difficult to imagine how very visually complex art works,

---


232 The type of medium used is very important to ordinary and professional practices regarding art. For instance, paintings are categorized by whether they are oil, watercolor, tempera, charcoal, or frescoes. Sculptures are categorized by whether they are marble, wood, or bronze. The medium also plays a role when we are trying to define art, and using the term “art” as an evaluative term. For instance, few people would refuse the label “art” to one of the bronze horses atop San Marco in Venice are works of art. But many people would hesitate to give the label “art” to an otherwise identical sculpture of a horse executed in butter at the Iowa State Fair.

such as Michelangelo’s *The Last Judgment* or Bosch’s *The Garden of Earthly Delights*, could have existed in all their glory solely within the minds of their respective artists. Surely the human imagination, even the imagination of a genius like Michelangelo, is not powerful enough to simultaneously visualize all of the details of figure, shape, color, shade, texture, movement, shadow, and the like that make up such art works. How then, can Croce and Collingwood argue that *The Last Judgment* existed fully and completely within Michelangelo’s mind?²³⁴

Croce and Collingwood’s theories of art face serious difficulties accounting for physical works of art like paintings and sculptures. But perhaps the idealist account of fictional works and fictional entities is still quite sound. After all fictional works and fictional entities are not concrete; the physical object hypothesis thoroughly fails when applied to literature.²³⁵ While Croce and Collingwood may have erred by postulating paintings and sculptures as mental objects, we can still hold to their views that literary works, and their attendant characters, are mental objects.

However an idealist account of literature is not immune to the problems faced by idealist accounts of painting and sculpture. According to Croce and Collingwood, fictional works and fictional entities, being art works, exist fully and completely formed within the minds of their authors. An author does not need to write, type, or publish anything in order to create a fictional work and fictional characters. The only purpose of creating a physical copy of a literary work is to communicate to other people the story and characters that already exist within the author’s mind. Now it is not difficult to imagine an author mentally creating a short story or poem in its entirety without having

²³⁴ This difficulty is discussed in Kemp, “The Croce-Collingwood Theory as Theory,” 184 and in Wollheim, *Art and its Objects*, 42.
²³⁵ Wollheim, *Art and its Objects*, 4-10.
to write anything down. Nor is it difficult to imagine an author mentally creating a simple fictional entity, one that possesses relatively few properties, without having to write anything down. But it is difficult to imagine how an author could create a very long and complicated work of literature like *War and Peace* or *Finnegans Wake* entirely within his mind. Such a feat seems to be beyond the capabilities of the human imagination.

There are much more serious objections to Croce and Collingwood’s theories of fictional realism. Even if we assume for the sake of argument that it is somehow possible for a very complicated fictional work like *Finnegans Wake* to exist entirely within the mind of an author, postulating fictional entities and fictional works as mental objects has unappealing and counter-intuitive consequences. This is surprising, given that Croce and Collingwood’s account of the author is refreshingly commonsensical compared to the author-as-discoverer view advocated by Meinong, Parsons, and Wolterstorff. Croce and Collingwood are correct to tie the existence of fictional entities and works to authors, but they have tied the knots too tightly.

Croce and Collingwood both agree that all works of art, including fictional works and fictional entities are purely mental entities. Being purely mental entities, fictional works and characters only exist within minds. Specifically, a fictional work or entity exists only within the mind of the author who created it. Hamlet exists only within the mind of Shakespeare. He does not exist in the spatio-temporal world of concrete objects, he does not exist in Plato’s heaven, and he is not to be found among Meinong’s infinite array of non-existent objects.
Since Hamlet existed in his entirety within Shakespeare’s mind, it would have been very easy for Shakespeare to “access” or to perceive Hamlet. In other words, Shakespeare would have been able to know what Hamlet was like, what properties Hamlet had, and such. This is not simply because Shakespeare created Hamlet, it is also because Hamlet is identical to a specific mental state of Shakespeare’s. However the contents of one’s own mind are essentially internal and private. My thoughts do not exist outside of my mind or independently of my mind. My thoughts “belong” to me and to me alone. No other person can have my thoughts. No other person can “access” my thoughts either. No other person can perceive my thoughts directly as I can, and no other person can know what my thoughts are like. At best, I can describe certain mental states of mine in detail to another person so he has an idea of what my mental states are like. But whatever I communicate will not enable the other person to directly perceive my thoughts. A description, even a very detailed description, of thought $x$ is not the same as thought $x$ itself just as a photo of the Eiffel Tower is not the Eiffel Tower itself.

Applied to the case of fiction works, this means that no other person besides Shakespeare can perceive *Hamlet* itself. *Hamlet* is identical to a mental state that Shakespeare had. Being a mental state, Hamlet is essentially internal and private to Shakespeare. Shakespeare wrote a script for *Hamlet* in order to describe and communicate this mental state to other people. We can perceive Shakespeare’s description of his mental state. But Shakespeare’s description of *Hamlet* is not identical to *Hamlet* itself. Likewise, all of the characters in *Hamlet*, like Hamlet and Ophelia, existed solely and irretrievably within Shakespeare’s mind. We can perceive
Shakespeare’s description of Hamlet and Ophelia via the script of *Hamlet* or a performance of *Hamlet*, but we still will not perceive Hamlet and Ophelia themselves.

Both Croce and Collingwood admit this disconcerting consequence of idealist fictional realism. A work of art remains within the mind of the artist. Spectators of the art work - the audience at a play, visitors at a museum, readers of a novel - perceive whatever physical object the artist has created in order to communicate the art work to others. By means of this physical object, spectators recreate their own personal “copies” of the art work.\textsuperscript{236} This is why Croce uses the term “physical stimulants of reproduction” to describe things like paintings, printed books, and musical scores.\textsuperscript{237} The task of an audience member of an art work is to recreate the artwork herself. And as Collingwood points out, the object that is re-created by an audience member is not identical to the object created by the artist.

Our own attitude towards Hamlet is that we do not know him, we imagine him. . . The Hamlet that we imagine is created by our act of imagining him; the Hamlet that Shakespeare imagined, by Shakespeare’s act of imagining him; and these two Hamlets, though they may resemble each other, are not identical. Whereas the London that I know and the London that you know are the same London, and this London does not depend for its existence on your personal acquaintance with it, nor yet on mine.\textsuperscript{238}

So when I read the script or attend a performance of *Hamlet*, I recreate anew the literary work *Hamlet* as well as all of the characters within it. My creations are not identical to those of Shakespeare. Certainly my creative activity is informed by Shakespeare’s, but my *Hamlet* is not his *Hamlet*. And both of these *Hamlets* depend on our respective


\textsuperscript{237} Croce, *The Aesthetic as the Science*, 108.

\textsuperscript{238} Collingwood, *Outlines of a Philosophy of Art*, 12.
mental activities; there is no one *Hamlet* that exists out in the world independently of my mind or Shakespeare’s mind.

Furthermore if you read the same script or attend the same performance of *Hamlet* with me, you too will recreate anew *Hamlet* and all of its characters. Your *Hamlet* will be similar to Shakespeare’s *Hamlet*, but it will not be identical to it. Your *Hamlet* will not be identical to my *Hamlet* either, since your *Hamlet* will be one of your mental states, and your mental states are never my mental states. For every reader of *Hamlet*, there will be a distinct Hamlet, a distinct Ophelia, and a distinct *Hamlet*. And all of these Hamlets, Ophelias, and *Hamlets*, will be distinct from the Hamlet, Ophelia, Polonius, and *Hamlet* created by Shakespeare himself.

This is a very odd state of affairs, but it is an inevitable result of Croce and Collingwood’s insistence that fictional works and entities are purely mental entities. I hesitate to call this consequence of idealist fictional realism “counter-intuitive” because I am not certain that ordinary intuition extends this far into matters of identity. But it certainly is an unusual and unappealing account of fictional works and fictional entities. It has always seemed to me that, for instance, Shakespeare created a literary work known as *Hamlet*. *Hamlet* is a contingent entity and exists solely due to Shakespeare’s creative activity, much of which was purely mental activity. But once created by Shakespeare, *Hamlet* exists and exists independently of Shakespeare’s mind. When I watch a performance of *Hamlet* I truly perceive *Hamlet*, not just a description of it provided by Shakespeare. I perceive the same *Hamlet* that Shakespeare created, I perceive the same *Hamlet* that audience members at the Globe perceived four hundred years ago, and I perceive the same *Hamlet* that the person sitting next to me in the theater perceives.
Likewise for the characters in *Hamlet*; the Ophelia created by Shakespeare, perceived by me, and perceived by you is one and the same Ophelia.

Meinongianism and Platonism err by severing the intuitive connection between an author and her fictional work and entities. Croce and Collingwood err by making this connection too tight. Fictional works and entities are indeed created by their authors. Much of the creative process involves purely mental activities on the part of the author. But once created, fictional works and entities must enjoy some level of independence from their authors. While fictional works and entities are not physical objects, they somehow need to exist “out” in the world, independently of the mind. Fictional works and entities require this ontological freedom so that two or more people can all perceive one and the same fictional work or entity.

An acceptable theory of fictional realism should be a *via media* between Meinongian-Platonist views and the Croce-Collingwood theory. Such a theory will grant fictional works and entities their contingent, created status and intimate connection to their authors. But at the same time, a theory of fictional realism will recognize that fictional works and entities enjoy a certain level of mind-independence. This is in addition to fulfilling the other criteria for a theory of fictional realism discussed in section 2.1.

2.8 Artifactualism

There is a theory of fictional realism that satisfies these requirements; it is known as artifactualism. Artifactualism is the claim that fictional entities, as well as fictional
works, are existent, abstract artifacts. The notion of an abstract artifact may strike many as bizarre; after all abstracta are usually conceived of as uncreated, necessary entities, while artifacts are taken to be concrete objects. But hopefully by this point it should be clear why it is necessary to postulate fictional entities as abstract artifacts. Whatever the merits of the physical object hypothesis for paintings and sculpture, it clearly fails with regards to fiction. Fictional entities and fictional works are not identical to any sort of physical object. They will be abstract objects. If we identify fictional entities and works with traditional abstracta like Meinongian non-existents and Platonic forms, then we cannot account for the intuitive created status of fictional entities and works. This is a fatal flaw in Meinongian, neo-Meinongian, and Platonist fictional realism. Yet while the author is an important figure in the creation and existence of a fictional work or fictional entity, we cannot bind fictional works and entities too closely to their authors. If we identify fictional works and entities with authorial mental states, as Croce and Collingwood do, then fictional entities and works will become hopelessly private and inaccessible.

By identifying fictional works and entities as abstract artifacts, all of these problems are solved. Fictional entities and works retain their non-physical status. They retain their created status as well, since artifacts are by definition created by human activity. But artifacts, unlike private mental states of an author or reader, are external, mind independent things. Multiple people can read about, think about, and speak about one and the same artifact.

Artifactualism conceived in this manner is a relatively recent form of fictional realism, receiving attention and development only within the last decade. But traces of
Artifactualism can be found as far back as the turn of the last century. Lipps’ “Further Considerations on ‘Empathy,’” which Meinong famously rebutted, contains what could be considered proto-artifactualist views of fictional entities. Regarding the character of Mephistus from Faust, Lipps writes:

I know that Goethe’s Mephistus is a mere poetic form, that there has never been a Mephistus, that therefore none of the words that Goethe makes him say have ever been uttered by him. Nevertheless, I can discuss how Mephistus replies to Faust or to the Lord in a certain passage. I can say that he ‘effectively’ responds in a certain way and not in another.

And it should be noted that with this I do not want to formulate a judgment either on my or on Goethe’s imaginative activity, but I formulate it on the person of Mephistus . . . for me or for my consciousness Mephistus’ answer does not belong to the past, but to the immediate present. Contrarily, Goethe’s imaginative activity is without doubt a past fact.

On the other hand, I do not speak at all of the historical Mephistus, but rather of Goethe’s or more precisely of the Mephistus of the story. But he has a peculiar form of existence. Without any doubt he was once called into existence by Goethe. But having been called into existence and having reached artistic representation in the words of literature, he has a type of reality; what he does and says is in a certain sense beyond any doubt a verifiable ‘fact.’

It would be undue speculation to conclude from this passage that Lipps himself was an artifactualist. But we do see some traces of the view here. First of all, Lipps speaks as a fictional realist, treating a fictional entity like Mephistus as an object that possesses properties - there are facts about what Mephistus is like. Lipps separates himself from the Meinongian and Platonist view of fictional entities by asserting that Goethe created Mephistus. However, Lipps appears to depart from the Croce-Collingwood theory of fiction and claim that Mephistus enjoys a certain amount of mind independence. A judgment that Mephistus possesses a specific property is not formulated “on my or on

239 Lipps, quoted in Raspa, “Fictional and Aesthetic Objects,” 52.
240 Lipps also differs from Meinong by asserting that Mephistus exists.
Goethe’s imaginative activity, but . . . on the person of Mephistus.” In other words, if I understand Lipps correctly, the claim that Mephistus possesses a specific property is not a claim about a mental state that Goethe had or a mental state that a reader has, but rather a claim about an object external to everyone’s mind, Mephistus.

The foundations of artifactualism are much more explicit in the work of the Polish philosopher Roman Ingarden (1893-1970). Ingarden’s work was primarily focused on metaphysics, specifically refuting his teacher Husserl’s transcendental idealism. But in his discussion of the idealism vs. realism controversy, Ingarden felt it very important to study the case of literature. Literary works and fictional characters, according to Ingarden, are excellent examples of mentally-dependent objects which provide a sharp contrast with objects in the “real world” which are not, contra Husserl, mentally-dependent.

While Ingarden holds fictional works and entities to be dependent upon mental activities, he does not conceive of them as mental entities in the conventional sense. Specifically, Ingarden rejects the Croce-Collingwood view of fictional works and entities as identical to mental states of their authors or readers. Ingarden points out that if a literary work like Hamlet is identical to a mental state had by Shakespeare, then it will remain essentially internal and private, as all mental states must be, and no one else will truly be able to perceive Hamlet. If Hamlet is identified with a mental state had by a reader or audience member, then there will be a separate and distinct Hamlet for every reader, resulting in the absurd conclusion that there are thousands of Hamlets.

---

241 Lipps, quoted in Raspa, “Fictional and Aesthetic Objects,” 52.
243 Ingarden, The Literary Work of Art, 15.
works must have a certain degree of mind-independence to account for our ordinary beliefs about them. But they cannot have this mind independence by being physical objects out in the real world; clearly literary works and fictional characters are not concrete. Nor are literary works and characters ideal, Platonic objects, since this would require them to have a timeless and immutable existence, which Ingarden thinks they clearly do not have. Furthermore, Platonic objects are uncreated and literary works clearly are created. While they are not identical to mental states, and are in some sense external to the mind, they still are dependent upon mental states and activities of their authors.

Fictional works and entities are, according to Ingarden, “purely intentional objects.” Purely intentional objects are contingent objects that depend upon specific human mental states and mental activities in order to exist. Such objects “draw the source of their existence and their essence directly from concrete acts of consciousness effected by an ego.” In other words, were it not for specific human authors, having specific mental states, and engaging in specific kinds of mental activities, there would be no *Pride and Prejudice*, *Hamlet*, or *Sherlock Holmes*. But such fictional works and entities are not identical to mental states had by Austen, Shakespeare, or Conan-Doyle. Instead fictional works, and their attendant characters, are a type of composite object. Literary works are composed of linguistic entities, of words and sentences. Ingarden actually defines four distinct linguistic “strata” that, like voices in a four-part choir,
combine in “polyphonic harmony” to form a literary work.\textsuperscript{247} Ingarden’s account of the strata is quite detailed, but it can be summed up by the claim that literary works are composed of sentences and depend upon these sentences in order to exist. If the sentence “Call me Ishmael,” the sentence “It was the devious-cruising Rachel, that in her retracing search after her missing children, only found another orphan,”\textsuperscript{248} and all the sentences in between these two had never been written, then the novel \textit{Moby Dick} would not exist, and nor would its fictional entities such as Ahab and the whale exist.

Sentences are also objects and have a particular mode of being which Ingarden calls “borrowed intentionality.”\textsuperscript{249} Objects with borrowed intentionality are initially created by human mental activity and depend upon consciousness in order to exist. “Sentences are not ideal objects that are totally independent of subjective acts in their existence and their properties. They are the products of these acts.”\textsuperscript{250} But once created these objects enjoy a certain amount of mind-independence. They can still exist even if their creator is not presently thinking about them. Borrowed intentionality allows “objects to free themselves, so to speak, from immediate contact with the acts of consciousness in the process of execution and thus to acquire a relative independence from the latter.”\textsuperscript{251} Objects with borrowed intentionality also become public objects, able to be accessed by multiple people.

These objects gain a certain advantage over primary purely intentional objects. for while the latter are “subjective” formulations, in the sense that in their

\textsuperscript{247} Ingarden’s account of language and its strata is quite complicated. The details of it are beyond the scope of this discussion but they can be found in \textit{The Literary Work of Art}, 29-117.
\textsuperscript{248} Herman Melville, \textit{Moby Dick} (New York: Penguin Putnam, 2000), 25, 638.
\textsuperscript{249} Ingarden, \textit{The Literary Work of Art}, 125-126.
\textsuperscript{251} Ingarden, \textit{The Literary Work of Art}, 126.
primariness they are directly accessible only to the one conscious subject who
effected the act that created them, and while in their necessary belonging to
concrete acts they cannot free themselves from these acts, the derived purely
intentional object can be intended or apprehended by various conscious subjects
as identically the same.\textsuperscript{252}

A sentence like “Call me Ishmael” is both mind dependent and independent. Were it not
for Melville engaging in a particular kind of creative, mental activity, this sentence never
would have existed. But having been created by Melville, this sentence can continue to
exist independently of Melville’s mind. While Melville had to have a thought along the
lines of “Call me Ishmael” to create this sentence, “Call me Ishmael” still exists even
after Melville’s thoughts turn to other matters. And even though Melville is the creator
of “Call me Ishmael,” he is not the only person who can truly access this sentence. The
sentence is a public and external object; all those who read the first three words of \textit{Moby Dick}
perceive one and the same sentence.\textsuperscript{253}

In virtue of the fact that they are constituted by sentences, literary works and
fictional characters will also enjoy this pleasant combination of mental dependence and
independence. \textit{Moby Dick}, Ahab, and the whale owe “the source of their existence and
their essence directly from concrete acts of consciousness,”\textsuperscript{254} engaged in by Melville.
But once created, \textit{Moby Dick}, Ahab, and the whale exist independently of Melville’s
conscious activity. Thousands of readers can perceive one and the same Ahab. Ahab
will continue to exist even if no one is presently thinking about him. As long as the

\textsuperscript{252} Ingarden, \textit{The Literary Work of Art}, 126.
\textsuperscript{253} Thomasson, \textit{Fiction and Metaphysics}, 22-23.
\textsuperscript{254} Ingarden, \textit{The Literary Work of Art}, 118.
sentences that constitute *Moby Dick* exist, this work and its characters continue to exist.\(^{255}\)

Under Ingarden’s account of literature, fictional works and fictional entities occupy a very unusual, “hybrid” category of objects. They are abstracta, but contingent. They are created by mental activities, but exist externally to the mind. Ingarden admits that fictional works and entities do not neatly fit into the traditional categories of spatio-temporal objects, abstract objects, and mental objects. “It is neither among physical, psychical, nor psycho-physical objects, that we find literary works.”\(^{256}\)

Ingarden’s suggestion that fictional entities are categorically idiosyncratic objects, inspired the work of Amie Thomasson who has provided the most detailed and developed account of artifactualism to date.\(^{257}\) Thomasson’s work, particularly her 1999 book *Fiction and Metaphysics*, can be credited with establishing artifactualism as a viable alternative to neo-Meinongian, Platonist, and idealist forms of fictional realism.\(^{258}\)

Thomasson begins with the observation that ordinary discourse and intuitions about fiction treat fictional entities as objects created by the actions of their authors.\(^{259}\) Sherlock Holmes exists only because he was created by Arthur Conan Doyle. If Conan Doyle had not engaged in activities of a certain sort (specifically writing the Sherlock Holmes stories) then Sherlock Holmes never would have existed. Ordinary intuition does

\(^{255}\) Thomasson, *Fiction and Metaphysics*, 23.


\(^{257}\) The term “artifactualism” is Thomasson’s.


\(^{259}\) Thomasson, *Fiction and Metaphysics*, 5-6.
not see fictional entities as eternal and necessary entities like Meinongianism and Platonism do, and ordinary intuition is correct on this point, according to Thomasson.

Fictional entities are essentially created objects. Just as a frying pan, a lighthouse, or a bicycle cannot exist unless created by a human or humans, Sherlock Holmes, Hamlet, and Jane Eyre cannot exist unless created by humans. Since fictional entities depend on human creative activity in order to come into existence, Thomasson argues that such entities are best categorized as artifacts. And while fictional entities are in the same category as frying pans and bicycles, there is an important difference between them and the latter. Fictional entities are not concrete, physical objects. They are not made of matter and do not occupy a definite position in space-time. Instead, fictional entities are abstract objects.

This is an unusual claim, to be sure. Abstract objects, traditionally conceived, are timeless entities which exist from eternity. Fictional entities come into existence at a specific point in time. For instance, “if someone contended that George Washington was a great fan of Sherlock Holmes, we might object that in Washington’s time there was no Sherlock Holmes - the Holmes character was not created until 1887.” Abstracta are also traditionally thought to be necessary entities. Fictional entities are contingent entities. They only exist in virtue of their authors engaging in specific sorts of creative activities; fictional entities do not have to exist. “At times we count our good luck that certain characters like Sherlock Holmes were created when, given a busier medical practice, Arthur Conan Doyle might never have created him.”

---

261 Thomasson, Fiction and Metaphysics, 5-6.
262 Thomasson, Fiction and Metaphysics, 6.
Fictional entities are a hybrid object of sorts: contingent abstracta, or abstract artifacts. Thomasson acknowledges that this “new” category of objects will raise many skeptical objections, but seeks to allay this by pointing out that many familiar entities, like marriages, contracts, laws, universities, and institutions also are abstract artifacts. For example, the speed limit on the New Jersey Turnpike is an abstract artifact. The speed limit is not any physical object. It is not identical to a printed copy of the law in the statehouse, or to the collection of road signs posted along the Turnpike, or to the Turnpike itself, or to any police officer that enforces the law, or to any paper speeding ticket, and so on. Still the speed limit exists and is some sort of thing. The speed limit is best seen as an abstract object. However, it is certainly not an abstract object like a Platonic form or number. The speed limit on the New Jersey Turnpike does not exist necessarily and timelessly. It came into existence at a particular point in time, and its existence is highly contingent. Lawmakers could have set a different speed limit for the Turnpike, or they could have decided to set no speed limit at all, or the Turnpike itself could have never been constructed, or the automobile could have never been invented, and so on. Thomasson points out that despite the fact that things like laws, institutions, marriages, and such straddle the traditional abstract-concrete divide, few philosophers

---


264 No one, not even a philosopher, would defend himself against a speeding ticket by claiming that since the speed limit is not to be found anywhere in space-time, ergo the speed limit does not exist.
have any qualms about admitting them into our ontology.\textsuperscript{265} Since fictional entities are in the same category it would make little sense to exclude them from our ontology.\textsuperscript{266}

Fictional entities are essentially created objects, dependent upon the activities of their authors in order to exist. Although fictional entities are dependent upon the creative activities of their authors, Thomasson does not identify such entities with specific mental states of their authors (or of their readers). Like Ingarden, Thomasson argues that once a fictional entity is created by an author’s mental activity, the fictional entity continues to exist independently of the author’s mind. Fictional entities do not need to be constantly thought about by their authors. Furthermore, fictional entities can continue to exist even after their authors have died. Hamlet still exists four hundred years after the death of Shakespeare.\textsuperscript{267} This makes sense given that fictional entities are artifacts. A conventional artifact, like a clay pot, requires a person to bring it into existence. But once brought into existence, the pot exists independently of the potter, and may in fact continue to exist for centuries after the potter has died.

Although fictional entities enjoy continued existence, sometimes over centuries, they still remain dependent objects. A fictional entity is brought into existence by its author, but it is sustained in existence not by its author but by the literary work in which it appears. Fictional entities are never presented to us by themselves; they are always found “within” some novel, short story, poem, drama, or similar literary work.\textsuperscript{268} Indeed, fictional entities depend upon fictional works in order to exist. If \textit{Hamlet} had never been

\begin{itemize}
\item \textsuperscript{265} Thomasson, \textit{Fiction and Metaphysics}, 13-14, 55, 73, 139, 144-146, 148-153.
\item \textsuperscript{266} Thomasson, \textit{Fiction and Metaphysics}, 139-153.
\item \textsuperscript{267} Thomasson, \textit{Fiction and Metaphysics}, 9.
\item \textsuperscript{268} Fictional entities, being abstracta, are of course not literally “within” anything. Such terminology is strictly metaphorical. See \textit{Fiction and Metaphysics}, 7-8.
\end{itemize}
written, Hamlet, Ophelia, and Polonius would never have existed. It is through creating a fictional work, in this case Hamlet, that an author like Shakespeare creates fictional entities. The creation of a fictional work is the specific kind of mental activity that an author must perform in order to create fictional entities. And once the fictional work has been written, its continued existence provides grounds for the continued existence of its characters. Hamlet and Ophelia exist to this day because *Hamlet* exists to this day, and Hamlet and Ophelia will continue to exist as long as *Hamlet* itself does.

Since the continued existence of a fictional work is required for the continued existence of fictional entities, naturally the question arises, what kind of object is a fictional work? According to Thomasson, literary works are the same kind of object as fictional characters, abstract artifacts. Fictional works are abstract entities, but they are contingent, temporal, and essentially created. They are brought into existence at a particularly point in time by the creative activities of their authors. Once created by an author, a fictional work enjoys existence independent of the mind, but its continued existence is still dependent on other factors.

While a fictional work is not identical to any particular physical copy of it or even to the class of all its physical copies, it does require the presence of at least one copy to remain in existence. This copy can be a printed paper copy, an electronic copy, an audio copy, or a memorized copy. Some fictional works, such as folktales, are never

---

written down or put into some other physical form, but they are preserved in existence due to mental “copies” in the memories of various people.\footnote{273}{Thomasson, Fiction and Metaphysics, 9.}

But more than physical copies is required to sustain a literary work in existence. All works of literature are composed in some human language - English, French, Russian, etc. Human languages can, and sometimes do, cease to exist. And when a language ceases to exist, Thomasson argues, all literary works in that language die along with it. This happens even physical copies of this tongue’s literary works remain. Such copies will be nothing “but some ink on paper.”\footnote{274}{Thomasson, Fiction and Metaphysics, 11.} So in order for a fictional work to enjoy continued existence, the language that the work is written in must continue to exist. And in order for a human language to continue to exist, there must be some people who have the ability to understand it, either in spoken or written form. Thus, what is actually required for the continued existence of a fictional work is at least one person who knows the language the work is written in and can understand the work. As Thomasson puts it:

A literary work as such can exist only as long as there are some individuals who have the language capacities and background assumptions they need to read and understand it. If all conscious agents are destroyed, then nothing is left of fictional works . . .\footnote{275}{Thomasson, Fiction and Metaphysics, 11.}

It is very easy, then, to conceive of scenarios where a fictional work is destroyed, never to be recovered. A fictional work can be lost forever if a Fahrenheit 451-esque situation occurred and all physical copies of the work were destroyed before any people had a chance to memorize the work. A fictional work can be lost forever if it is a folktale persevered only through oral tradition, and everyone in the cultural group who knows the
tale dies without passing it on to the next generation. A fictional work can be lost forever if it is composed in a language like Etruscan that no living person can speak or read.

And since fictional entities depend upon fictional works to continue to exist, fictional entities can likewise be destroyed by such conditions. If every physical, electronic, and audio copy of *Pride and Prejudice* were to be destroyed and every person who had memorized *Pride and Prejudice* killed, then Darcy and Lydia Bennett would no longer exist. The characters of whatever fictional works the Etruscans may have composed no longer exist either.

This is atypical of abstract objects. Platonic forms, numbers, Meinongian non-existents, and the like cannot be destroyed. They always exist\(^\text{276}\) no matter what states of affairs obtain in the world of contingent, spatio-temporal objects. But the impermanence displayed by fictional works and entities is quite typical of artifacts. Artifacts can be and routinely are destroyed; a vase can be broken into thousands of pieces, a house can be burnt to ashes, a child’s shoebox diorama can be eaten by a dog. A fictional work or entity that is lost and gone forever is the abstract equivalent of a burnt house or eaten diorama.\(^\text{277}\)

Having established that fictional entities are abstract artifacts, Thomasson turns her attention to what fictional entities are like. She notes that folk-level discussions of fiction and discussions done at the level of literary scholarship both speak of fictional entities as having properties: Hamlet is a prince, Hamlet is Danish, Hamlet was created by Shakespeare, and the like. Fictional entities do indeed have properties, according to

\(^{276}\) Or are always objects, in the case of Meinongian non-existents.

\(^{277}\) Thomasson, *Fiction and Metaphysics*, 10; Thomasson, “The Ontology of Art,” 79.
Thomasson. However, it is very important to distinguish two levels of discourse about fictional entities. The properties that fictional entities are said to have can be divided into two classes corresponding to these levels of discourse.  

Sometimes when we speak about fictional entities we engage in what Thomasson calls “fictional discourse.” A statement about a fictional entity qualifies as fictional discourse if it could be prefixed with “according to the story” without becoming false. For example a statement like “Hamlet is a prince” is part of fictional discourse because, once modified as “according to the story, Hamlet is a prince,” it is still a true sentence. Fictional discourse, Thomasson admits, always involves a certain level of pretense. When we engage in fictional discourse, “we pretend that what the story says is true; for example, that it describes not mere fictional characters but instead actual people whom we could praise, criticize, or psychoanalyze.” But throughout this discourse we know that the story is not true, that the characters are not real people, and that the events depicted did not actually occur.

At other times we speak about fictional entities in what Thomasson calls “real context” or what I will call “external discourse.” External discourse is discourse about fictional entities that takes place “externally” from the story. It is the discourse of the critic, the scholar, and the philosopher. Such discourse does not engage in pretense but treats fictional entities as fictional. A statement about a fictional entity belongs to external discourse if it cannot be prefixed with “according to the story” without becoming false. For example, a statement like “Hamlet was created by Shakespeare” belongs to

---

279 Thomasson, Fiction and Metaphysics, 105.
280 Thomasson, Fiction and Metaphysics, 105.
281 Thomasson, Fiction and Metaphysics, 106.
external discourse because if it is prefixed to “according to the story, Hamlet was created by Shakespeare” it becomes false. According to the story, Hamlet was most certainly not created by Shakespeare; according to the story he was created by his parents Gertrude and the murdered king.

Fictional entities only literally or genuinely possess the properties predicated of them in external discourse. Hamlet literally is an abstract artifact and literally was created by Shakespeare. Hamlet has the property of being an abstract artifact in the same straightforward manner as Las Meninas has the property of being a concrete artifact, and Amy Carter has the property of being a human. Hamlet has the property of being created by Shakespeare just as Las Meninas has the property of being created by Velazquez, and Amy Carter has the property of being created by Jimmy and Rosalyn Carter.  

Meanwhile fictional entities do not literally have the properties predicated of them in fictional discourse. Hamlet is not literally a prince; Hamlet was not literally created by Gertrude and the murdered king. Instead Hamlet has the properties of being a prince according to the story, and being created by Gertrude and the murdered king according to the story. But the property of being a prince according to the story is not identical to the property of being a prince. In other words, Hamlet is, in a sense, a prince, but he is not a prince in the same manner that Charles Windsor is a prince.

Still fictional entities possess all the properties attributed to them by their stories as long as these predications are implicitly prefixed with “according to the story.” Fictional entities also possess properties that are implied but not explicitly attributed to

283 Thomasson, Fiction and Metaphysics, 113.
them by their stories, as long as these predications are similarly prefixed.\textsuperscript{284} Although \textit{Pride and Prejudice} never explicitly states that Lydia Bennett has feet, the story implies this, so Lydia Bennett has the property of having feet according to the story (but not, of course, the property of having feet).

The distinction between fictional and external discourse allows Thomasson to avoid the anti-realist charges that fictional entities sometimes violate the law of non-contradiction. Imagine, for instance, that one were to write a story about the adventures of Meinong’s infamous round square. Anti-realists would claim that this story has, in virtue of creating a fictional entity, created an object. This object, the round square, is an impossible object because it has the property of being both round and square. But Thomasson would argue that a statement about this character like “the round square is both round and square” is the kind of statement that would occur in fictional discourse. Such a statement describes what is the case within the world of the story. Since this predicative statement occurs within fictional discourse, it must be implicitly prefixed with according to the story. Consequently, the round square will have the property of being both round and square \textit{according to the story}. The property of being both round and square according to the story is not the same as the property of being both round and square. And while it is impossible for an object to have the property of being both round and square, there is nothing particularly problematic about an object having the property of being both round and square according to some story.\textsuperscript{285}

Similar principles apply to the anti-realist charge that all fictional entities are intolerably incomplete. Hamlet, for example, is incomplete with respect to the property

\textsuperscript{284} Thomasson, \textit{Fiction and Metaphysics}, 107.
\textsuperscript{285} Thomasson, \textit{Fiction and Metaphysics}, 107.
of having bloodtype A. The story *Hamlet* does not explicitly attribute the property of having bloodtype A to Hamlet or explicitly state that Hamlet lacks the property of having bloodtype A. Hamlet’s specific bloodtype is not implied by anything in the story either, as Lydia Bennett’s possession of feet is implied by *Pride and Prejudice*. But according to Thomasson, Hamlet is not incomplete with respect to having or lacking bloodtype A. She notes

> Because these claims seem to treat Hamlet as a human being who has blood, it seems most likely that such statements would be made in a fictional context, in which case they should be read as “according to *Hamlet*, Hamlet is of bloodtype A” and “according to *Hamlet*, it is not the case that Hamlet is of bloodtype A.” Both of these are false, because the story does not mention anything about it. But because these sentences do not describe real properties of Hamlet, but only how he is said to be according to the story, we need not infer on that basis that Hamlet is incomplete with respect to the property being of bloodtype A.286

Outside of the story, Hamlet qua fictional entity is not incomplete at all. He is complete with respect to having bloodtype A; he definitively lacks this property being an abstract artifact with no blood.287 All other fictional entities are similarly complete. Fictional entities only appear to be intolerably incomplete within their stories. But fictional entities never literally have the properties given to them by their stories, including the property of being incomplete. Being incomplete according to a story is not the same property as being incomplete. And while no object can be incomplete, an object can be incomplete according to a story without raising any serious metaphysical problems.

---

2.9 Concluding Remarks

Artifactualism is a satisfactory middle course between the Scylla of Meinongian and Platonist fictional realism and the Charybdis of Croce-Collingwood idealism. The greatest flaw of Meinongian, neo-Meinongian, and Platonist theories of fictional realism is their extremely counter-intuitive view of authors as discoverers, not creators, of fictional works and entities. Artifactualism returns to an intuitive picture of the author as a creator of fictional works and entities. Artifactualism also accounts for how fictional works and entities, despite being non-physical objects themselves, seemingly depend on spatio-temporal objects like printed books and people for continued existence. Intuitions about this issue are particularly troublesome for Meinongian and Platonist fictional realism, as I pointed out in section 2.5. And although fictional works and entities are dependent objects under artifactualism, they enjoy a fair amount of mind independence. They are not mental entities, as Croce and Collingwood assert, and are not subject to the difficulties that follow from Croce and Collingwood’s views.

The greatest drawback to artifactualism is the discomfort that some philosophers will have with Thomasson’s “new” category of abstract artifacts. Undoubtedly some metaphysicians will find the notion of contingent abstracta incoherent. I believe this objection is surmountable; Thomasson more than adequately responds to it by showing that every one already accepts contingent abstracta into our ontology by accepting the existence of laws, institutions, contracts, and the like. Nonetheless it is beyond the scope of this dissertation to offer a sustained defense of artifactualism against this objection.

Van Inwagen raises this objection in “Existence, Ontological Commitment, and Fictional Entities,” 153-155, although he does not seem strongly committed to it.
Rather I hope I have shown why artifactualism is preferable to any other form of fictional realism.
3

Artifactualism and Interpretation

The previous two chapters have been devoted to careful study of the ontology of fictional characters. I have claimed that realism about fictional characters is a plausible position and that Amie Thomasson’s theory of artifactualism is the most plausible form of fictional realism. But as I alluded to at the end of Chapter 2, Thomasson’s account of the ontology of fictional characters is only coherent within her overall theory of the ontology of fictional works. The artifactualist account of literary works is quite interesting in its own right. So I will now switch gears somewhat and spend the remainder of this project focusing upon the ontology of literary works.

Here in this chapter I would like to do several things. First I will provide a more detailed overview of the artifactualist theory of literary works as articulated by Thomasson. For the most part, I agree with Thomasson’s formulation of artifactualism, but I will attempt to develop certain parts of the theory further than she does. Specifically, Thomasson does not say much about the role of interpretation in sustaining literary works in existence; I will argue that, assuming artifactualism is true, in certain situations the misinterpretation of a work of literature can result in its destruction.

Additionally I will discuss one of the major benefits of accepting this revised form of artifactualism. By viewing a literary work as dependent upon – but not identical
to – a proper interpretation of it, artifactualism presents an attractive alternative to two problem-fraught work identities theories: textualism and constructivism.

3.1 The Artifactualist Account of Literary Works

This section will summarize the artifactualist account of literary works as it is presented by Thomasson. I recognize that many aspects of Thomasson’s version of artifactualism are not without controversy, but it is beyond the scope of my project here to provide detailed arguments for artifactualism and to respond to specific objections to it. Instead I will largely take Thomasson’s account of artifactualism for granted, as I summarize the theory and consider its more interesting implications.

Artifactualism holds that literary works, like fictional characters, are abstract artifacts. Contrary to Meinongian and Platonist accounts of literature, artifactualism asserts that literary works are contingent entities. A literary work comes into existence when its author engages in specific creative activities while holding specific intentional states. But once it has been created, a literary work no longer depends on its author to enjoy continued existence. Literary works are not identical to authorial mental states, as idealism claims. Rather literary works can continue to exist long after their authors have died. The continued existence of the work depends on the existence of a combination of spatio-temporal objects and intentional states.

A central claim of artifactualism is that literary works are created at a specific point in time. Literary works do not exist from eternity in Plato’s heaven; literary works are not discovered among the infinite realm of Meinongian non-existents. Rather at a
specific point in time, an author or group of authors engages in a particular sort of activity which results in an entirely new object that did not exist beforehand.\(^\text{289}\)

Thomasson believes that the artifactualist account of literary works as created objects is most consonant with everyday and professional practices surrounding literature.

As ordinarily treated in critical discourse, a literary work is not an abstract sequence of words or concepts waiting to be discovered by instead is the creation of a particular individual or group at a particular time in particular social and historical circumstances. Thus, as with characters, it seems that literary works must be created by an author or authors at a certain time in order to come into existence.\(^\text{290}\)

According to artifactualism, literary works also are artifacts. By definition they are the product of human activity. Ontologically speaking, the author of a literary work has done the same thing as a carpenter, potter, or baker. All of these artisans bring new objects into existence, objects which did not exist beforehand. However carpenters, potters, bakers, and other artisans who bring concrete objects into existence do so by engaging in a physical process - usually involving giving some form, shape, or arrangement to some pre-existing physical matter. The creative activity of an author is quite different, since literary works are abstracta. When an author creates, or \textit{composes}, a literary work he does so by creating a \textit{text}.\(^\text{291}\) A text is “a sequence of symbols in a language or languages.”\(^\text{292}\) Usually the composition of a text results in the creation of a


\(^{290}\) Thomasson, \textit{Fiction and Metaphysics}, 8.

\(^{291}\) Thomasson, \textit{Fiction and Metaphysics}, 64.

\(^{292}\) Thomasson, \textit{Fiction and Metaphysics}, 64.
physical object - ink marks on a piece of paper or data encoded into a computer’s hard drive.

The composition of a text must be accompanied by specific authorial mental states; the author must intend to compose a literary work as he creates his text. Artifactualism rules out “Swamp-Pride and Prejudice” and similar cases. The result of this creative process is a literary work - a short story, novel, poem, or drama - an entity that has “certain aesthetic and artistic qualities and ordinarily telling a tale concerning various characters and events.” According to Thomasson, a literary work and its text are distinct entities. She rejects textualism, the claim that a literary work is identical to its text. This point will be very important later on.

Literary works are created at a specific point in time. However, it is usually impossible to state exactly when a given literary work has come into existence; literary works are susceptible to sorites-style paradoxes. For instance it is safe to say that when Herman Melville completed the final draft of Moby Dick, which he submitted to the publisher in 1851, Moby Dick had come into existence. But it is certainly not unreasonable to believe that Moby Dick existed before that moment. Perhaps before Melville finished the final draft, he went through the text and corrected spelling and grammatical errors. Or perhaps Melville decided, at the last minute, to change the name of his character “Starling” to “Starbuck.” It seems likely that Moby Dick already existed, even in its un-proofread-Starling-not-Starbuck state. Note that by making such minor revisions, Melville would have brought an entirely new text into existence yet he would

---

294 Thomasson, Fiction and Metaphysics, 64.
295 Thomasson, Fiction and Metaphysics, 64-66.
not have created a new literary work, given artifactualism’s rejection of textualism. But what if Melville had made more substantial revisions to *Moby Dick* before submitting it to the publisher? Suppose the tangential chapters that make up Ishmael’s account of whale taxonomy were a last minute addition. Were those chapters additions to an already existing *Moby Dick*, or was the composition of those chapters the crucial last step in a causal process that resulted in the creation of *Moby Dick*? Eventually this line of questioning leads one to ask how much of *Moby Dick*, how many of its chapters, its pages, had to have been written before *Moby Dick* came into existence. Clearly if Melville had stopped after writing “Call me Ishmael” he would not have created *Moby Dick*; he would not have created a literary work at all, just a sentence. If Melville had only written the first three chapters of *Moby Dick*, he would not have created *Moby Dick*, although he would have created a literary work, a short story about an uncomfortable overnight visit to Nantucket. But what happened when Melville had finished half of *Moby Dick*, or two-thirds of *Moby Dick*, or ninety-nine percent of *Moby Dick*? At one point in time did this new entity, that did not exist beforehand, come into existence?

Such questions are practically unanswerable since we do not possess the philosophical tools, specifically detailed identity conditions for literary works, to resolve the issue.\textsuperscript{296} However I do not believe that this vagueness presents an insurmountable problem for artifactualism. It is not necessary for artifactualists to establish at exactly which moment in time a literary work came into existence; that is, it is not necessary to establish that *Moby Dick* came into existence at 9:34 a.m. on September 22, 1851. What

\textsuperscript{296} Thomasson herself admits that artifactualism is unequipped to answer the questions raised by sorites-style paradoxes surrounding literary works. See “Debates About the Ontology of Art: What are We Doing Here?” 249-250; *Fiction and Metaphysics*, 110, 163 n. 25; “The Ontology of Art and Knowledge in Aesthetics,” 221, 227-228.
is necessary for artifactualists to establish is that *Moby Dick* came into existence at some specific point in time and that before this point in time, the novel *did not exist*. *Moby Dick* came into existence because at a specific moment in time, Melville intentionally engaged in a specific sort of activity, composition. Melville did not “discover” *Moby Dick*; he did not “select” *Moby Dick* from an infinite realm of Platonic entities or Meinongian non-existsents. And had Melville not created *Moby Dick*, *Moby Dick* would not exist. Had Melville never been born, *Moby Dick* would not exist. Every literary work depends upon its author in order to bring it into existence.\(^{297}\)

Not only does every literary work depend upon an author to bring it into existence, each work depends on its specific author in order to bring it into existence. Thomasson argues that all literary works have their authors essentially.\(^{298}\) Being created by Shakespeare is an essential property of *Hamlet*. Any literary work that lacks the property of being created by Shakespeare cannot be *Hamlet* even if this work is identical to Shakespeare’s *Hamlet* in all other respects. In fact, if by some incredible coincidence two separate authors compose two perfectly word-for-word identical texts, the result will be two numerically distinct literary works.\(^{299}\) This situation is notably illustrated in Borges’ short story “Pierre Menard, Author of the *Quixote*” in which a 20\(^{\text{th}}\) century Frenchman named Pierre Menard composes a perfect word-for-word replica of Cervantes’ *Don Quixote*.\(^{300}\) Menard’s *Don Quixote* has the same text as Cervantes’ *Don Quixote* yet Menard’s *Don Quixote* is a distinct literary work, due to its distinct author.


\(^{298}\) Thomasson, *Fiction and Metaphysics*, 8-9, 36, 132.

\(^{299}\) Thomasson, *Fiction and Metaphysics*, 8, 56-57.

And had Cervantes never been born, *Don Quixote* never would have existed. Of course another writer, whether in the 17th century or the 20th century, might have composed a novel entitled *Don Quixote* with the same text as Cervantes’ but this novel would not be identical to the *Don Quixote* that we read today but rather a completely different literary work.

Thomasson points out that once a literary work has been created by its author, it usually persists through time. *Jane Eyre* was created by Bronte in 1847, but it existed in 1947, it exists today in 2011, and it will continue to exist well into the foreseeable future. Literary works depend upon their specific authors to come into existence, but they clearly do not depend upon their authors to remain in existence. In most cases, literary works continue to exist long after their creators have died, just as sculptures and paintings do. As far as their continued existence is concerned, literary works are multiply dependent entities. They depend upon a combination of spatio-temporal entities and specific mental states. A literary work itself is not concrete, but it cannot persist through time unless a particular concrete object – at least one copy of its text – persists through time. A literary work itself is not a mental state held by an author or reader, but it cannot exist at a given moment in time unless a person with certain mental states – the language capacities and appropriate “background assumptions” required to read and understand the text – exists at that same given moment.

While a literary work is not identical to any particular physical copy of it or even to the class of all its physical copies, it does require the presence of at least one copy to

---

301 Thomasson, *Fiction and Metaphysics*, 11-12.
303 The notion that a literary work is identical to some spatio-temporal object is known as the physical object hypothesis. This hypothesis is widely discussed but not, to my knowledge, actually advocated by
remain in existence. In order for *Pride and Prejudice* to exist at time \( t \), at least one physical copy of *Pride and Prejudice* must exist at \( t \). This copy can be my copy, or the copy in the Rutgers library, or the original manuscript produced by Austen - it does not matter which copy exists as long as a copy exists. Furthermore this copy need not be a bound and printed book. An electronic copy or an audio copy of *Pride and Prejudice* will suffice. In fact if a savant were to memorize *Pride and Prejudice* verbatim, as in Ray Bradbury’s novel *Fahrenheit 451*, this memory would also qualify as a copy of *Pride and Prejudice*. After all, some fictional works, such as folktales, are never written down or put into some other physical form, but they are preserved in existence due to mental “copies” in the memories of various people.

A literary work also requires the presence of at least one person who can read it to remain in existence. In order for *Pride and Prejudice* to exist at \( t \), at least one person who can read and understand *Pride and Prejudice* must exist at \( t \). Thomasson points out that this person need not be anyone in particular - I can be the potential reader, you can be the potential reader, Eli Manning can be the potential reader, and so on. As long as there is at least one person who can read *Pride and Prejudice*, it does not matter who that person is.

In order to qualify as a potential reader of a literary work, a person must possess “the right language capabilities and background assumptions to read and understand the

---

305 Thomasson, *Fiction and Metaphysics*, 11.
306 Thomasson, *Fiction and Metaphysics*, 9, 11-12.
literary work.”\textsuperscript{308} The first part of this requirement is relatively simple. To possess the right language capabilities is simply to possess the ability to read texts written in a particular language. The language capability required by \textit{Pride and Prejudice} is the ability to read English, the language capability required by the \textit{Aeneid} is the ability to read Latin, and so on. The second part of this requirement is more intriguing. Thomasson claims that in order for a literary work to persist through time there must be at least one person who is fluent in the work’s language and who possesses the appropriate “background assumptions to read and understand the literary work.”\textsuperscript{309} In this case, understanding the literary work is something above and beyond being able to understand the language the work is written in.\textsuperscript{310}

Because it is so ontologically important for potential readers of a given literary work to have the “appropriate background assumptions” about the work, I must discuss in more detail what is meant by this terminology. The notion that a literary work depends, in part, on there being at least one reader able to properly understand it, is found in Thomasson’s \textit{Fiction and Metaphysics}. Unfortunately Thomasson does not give a detailed account of what proper background assumptions consist of, but what she does say about the issue provides a foundation on which to construct a more thorough theory.

3.2 Literary Interpretation

One plausible candidate for the kind of background knowledge and assumptions needed to understand a work is the sort of inference every reader makes about what is

\textsuperscript{308} Thomasson, \textit{Fiction and Metaphysics}, 65, see also 11-12.
\textsuperscript{309} Thomasson, \textit{Fiction and Metaphysics}, 65, see also 11-12.
\textsuperscript{310} Thomasson, \textit{Fiction and Metaphysics}, 11-12, 65-66.
true according to a story. For example, if one reads that a character dances, one automatically assumes that the character has legs (or, strictly speaking, has legs according to the story), even if the text of the story never explicitly states something along the lines of “Elizabeth has legs.” Most readers understand that the description of Elizabeth as dancing implies that she has legs because most readers are in possession of various background beliefs about dancing – humans typically have legs, dancing requires movement of the legs, and if a human is able to dance without legs this is an unusual and remarkable occurrence that would certainly have been noted by the author. Parsons notes how important such inferences are for understanding fiction:

Any reader who failed to extrapolate far beyond the printed word would be quite unable to understand the story. This is typically not a conscious process; we just instinctively assume all sorts of things on the basis of the given information, just as we do in perception. We read “Commander Roderick Blaine looked frantically around the bridge . . .” and we immediately assume, without conscious reasoning, that Blaine is named ‘Blaine’, that he is a commander, that he is male, that his eyes are open, that he is located in a space more than one cubic inch in volume, that he is upset . . . Any of these inferences may be revised in the light of further information (e.g., if we read that his eyelids had been replaced with some transparent material, and were now fastened shut).  

While background assumptions of this sort are necessary for understanding literary works, it seems that a reader must possess far more besides such simple inferences in order to understand most works of literature. I believe that Thomasson’s claim that a literary work depends up the existence of at least one person with the language abilities and proper background assumptions to read and understand it is best understood as the claim that a literary work depends upon the existence of at least one person who can properly interpret it.

311 Parsons, Nonexistent Objects, 176-177.
It is very instructive to observe how Thomasson draws a distinction between having the language background required to read a text and understanding a text. Frequently these abilities are one and the same; for simple and short texts like shopping lists, how-to manuals, and street signs any reader who is competent in the texts’ language can understand them. This is because such texts have no meaning beyond what is sometimes called their literal meaning. All that is required to grasp the literal meaning of a text is fluency in the text’s language.

In other cases, fluency in the language of a text is not sufficient for understanding the text. Consider this Latin sentence from Horace’s third satire. Regarding an attention-hungry singer, Horace writes, *si conlibuisset, ab ovo usque ad mala citaret “io Bacche!”*312 A competent speaker or reader of Latin should have no difficulty reading this sentence and ascertaining its literal meaning. Literally translated, this sentence means “if he himself was disposed, he would chant “Io Bacche” from the egg to the apples.” This is an odd thing to say. But this sentence has a non-literal meaning in addition to its literal meaning. The non-literal meaning of the sentence is quite different from the literal meaning and actually makes more sense. The Latin phrase *ab ovo usque ad mala* literally means “from the egg to the apples” but it does not really mean that. *Ab ovo usque mala* is an idiom which means “from the very beginning to the very end.” It is a reference to the grand multi-course banquets enjoyed in ancient Rome at which eggs were served as the very first course and apples were served as the very last course. Thus what Horace actually means to say is something along the lines of “if he himself was disposed, he would chant ‘Io Bacche’ from the very beginning of the evening to the very end.” If a

---

reader does not recognize the idiomatic character of ab ovo usque mala, then she will not be able to understand Horace’s text, even if she is perfectly fluent in Latin. She will know the literal meaning of the text but not its non-literal meaning. If she does not know this non-literal meaning of the text then she cannot be said to have understood the text.

Multiplicity of meaning – a tension between a text’s literal meaning and its implicit, hidden, or figurative meaning – is a feature of many works. Religious texts are generally thought to contain meanings beyond the literal; in fact, the earliest philosophical treatment of multiplicity of meaning concerns scripture, with the works of Philo of Alexandria, Augustine, and Averroes being most notable.313 Most works of literature also feature implicit, hidden, or figurative meanings in addition to their literal meanings. For instance, Animal Farm is a satire of Stalinism, although the text never says “this is all a satire of Stalinism, and Snowball is supposed to be Trotsky.” The Stranger has an existentialist theme, although the text never says, “Sartre is right, existence precedes essence.” Now in order to understand Animal Farm, a reader must understand – among other things – that it is a satire of Stalinism. I believe this is what Thomasson has in mind when she speaks of a reader who possesses the proper “background assumptions to read and understand the literary work.”314 Those “background assumptions” enable the reader to draw out the implicit, non-literal meaning of Animal Farm.

Now there is a very specific term for the drawing out the implicit meaning of a literary work – interpretation. Familiar synonyms of interpretation are criticism and

---

314 Thomasson, Fiction and Metaphysics, 65.
**Hermeneutics.** An alternative and more graceful way of saying a reader must know the implicit, hidden, or figurative meaning of a literary work in order to understand it is to say that a reader must properly *interpret* the literary work in order to understand it.

Although interpretation involves many other activities besides drawing out a text’s implicit meaning, this is one of its primary goals. \(^{315}\) “What does the literary interpreter do?” asks Monroe Beardsley, “He tells us what a literary work means.” \(^{316}\) A great deal of interpretation is concerned with the hidden meanings of works, meanings that are not literally stated in the text, meanings that are not always readily apparent to the reader.

The term interpretation implies that there is a puzzle of sorts to be solved. \(^{317}\) For this reason it sounds odd, humorous even, to speak of *interpreting* very simple and easily understood texts like a children’s story or a take-out menu. \(^{318}\)

But while much of interpretation is involved with discovering the non-literal meanings of literary works, it would be a mistake to see interpretation as limited solely to this activity. The “meaning” of a literary work is a rather flexible term; it involves many

---


\(^{316}\) Beardsley, “The Authority of the Text,” 24-25.

\(^{317}\) Iseminger, introduction, 1.

\(^{318}\) For a particularly amusing parody of this see, “Grad Student Deconstructs Take-Out Menu,” *The Onion*, July 24, 2002, accessed March 2, 2011, http://www.theonion.com/articles/grad-student-deconstructs-takeout-menu.85/. The graduate student laments, “I just wanted to order some food from Burrito Bandito. Next thing I know, I’m analyzing the menu’s content as a text, or ‘text,’ subjecting it to a rigorous critical reevaluation informed by Derrida, De Man, etc., as a construct or ‘construct,’ made up of multi-varied and, in fact, often self-contradictory messages, or ‘meanings.’”
other aspects of the work besides the explicit or implicit meaning of its words. As Noel Carroll writes:

... Literary meaning - that is, the object of literary interpretation - need not be concerned solely with the meaning of word sequences ... Literary interpretation may ask questions about the point of constructing a character in this way or that way and thus may investigate the representation of illocutionary acts in a text in terms of the contribution it makes to the point of the character as an element in the overall design of the work.319

The task of interpretation can be quite broad. Many philosophers take a good deal of interpretation to be a careful analysis and description of a literary work or certain parts of the work.320 In Hamlet and the Philosophy of Literary Criticism, Morris Weitz discusses the numerous tasks of a literary interpreter, in this case one who interprets Hamlet.

Among the descriptions are those of the characters, plot, dialogue, versification, images, metaphors, theme, symbolism, rituals, theatrical effects (e.g., cannon shot), and the sources and environment of the play. These descriptions may be individual or general: for example, descriptions of Hamlet’s or Claudius’ traits, or a description of the pervasive traits of all the characters.

The interpretations and explanations range from explanations of one item - for example, the imagery in relation to the whole play - to explanations of the whole play; often the explanations function as “readings” of the play in which “the meaning” of the play is made explicit in terms of what is “central” in it.321

321 Weitz, Hamlet and the Philosophy of Literary Criticism, 204-205.
Interpretation involves far more than simply explaining the meaning – explicit or implicit – of words and phrases in a literary work. Interpretation involves carefully analyzing and describing literary works and parts of literary works, such as their characters. It involves describing how certain parts of a work are related to other parts. It involves discovering and calling attention to any symbolism, allusions, metaphors, paradoxes, and imagery found in a literary work.\(^{322}\) It involves determining whether a work is allegorical, satirical, ironic, or transgressive.\(^{323}\) It involves discovering if the work is making an overall point about humanity, philosophy, religion, or society.\(^{324}\) Not all of these elements can be easily classified as “the meaning” of a literary work. According to Stein Haugom Olsen:

For the elements which can be understood as constituting a literary work, only a very few - for example, metaphors, symbols - can be naturally said to have meaning. It is . . . odd to talk about the meaning of characters, situations, actions, setting, rhymes, rhythm, stanza form, and so forth . . . to construe these diverse relationships and functions as a “juxtaposition” of elements which “interact” to create a “tensive meaning” is to stretch the concepts of “juxtaposition,” “interactions,” and “meaning” beyond the limits of their usefulness.\(^{325}\)


All of these other aspects, features, or properties of literary works are valid objects for interpretation. In fact, I would classify the explicit and implicit meanings of a literary work as properties of that work. Interpretation, then, is not best defined as discovering what a work means, as Beardsley claims. Rather interpretation involves discovering, calling attention to, or carefully analyzing all kind of features, parts, and properties of a literary work, including its meaning. A good working definition of interpretation is that it is the activity of saying something about a work’s meaning and significance. Given the multi-faceted nature of interpretation, perhaps the broadest and most general definition is best.

Hopefully this broad definition of interpretation underscores the commonplace and everyday character of this activity. Interpretation is often thought of as the domain of professional scholars and critics of literature, but it is performed by ordinary readers as well. The high school student who reads *Hamlet* and decides that Hamlet is insane, the beachgoer who notes the allusions to historical and religious figures in *The Da Vinci Code*, and the woefully uneducated reader who takes *Animal Farm* to be nothing more than a weird children’s story are all engaging in acts of interpretation. Interpretations need not be original, directed towards literature of consequence, or even correct in order to count as interpretations.

Now for any given literary work, certain parts of its meaning and significance will be more important than others for a reader to grasp if he is to understand the work. For example, it probably is not necessary for a reader to carefully note and analyze the colors of the cars driven by characters in *The Great Gatsby* in order for him to understand *The Great Gatsby*. This is not to say that an analysis of the car colors in *The Great Gatsby* is
not interpretation; it is because it deals with an aspect of *The Great Gatsby*’s meaning and significance. Rather this element of *The Great Gatsby* does not appear to be essential to a competent reading of the work. A reader who overlooked the car colors *The Great Gatsby* could still be said to have understood the novel.

On the other hand, there are many parts of a work’s meaning and significance that will be quite central and important for any reader who wishes to understand the work. The way that the characters and situations in *Absalom! Absalom!* function as metaphors for the Old South seems to be a very important property of that novel. A reader who failed to notice this aspect of *Absalom! Absalom!* would not have understood (or at least not fully understood) the work. The moral tension found in *Les Miserables* is an essential property of that novel. A reader who failed to notice this would not have understood *Les Miserables*. The allusions to Richardson’s *Pamela* found in Fielding’s parody novel *Shamela* are central properties of *Shamela*. A reader who was ignorant of these allusions would have understood *Shamela*. *Animal Farm*’s being a satire of Stalinism is an essential property of that novel. A reader who did not realize *Animal Farm* was satirical would not have understood it. And so on.

In order to understand a literary work, a reader must recognize and acknowledge the most important aspects of the work’s meaning and significance. Since interpretation is simply saying something about the meaning and significance of a literary work, another way to state this is that in order to understand a literary work, a reader must interpret the work properly. He must recognize the presence of the work’s most essential features.
Thomasson’s version of artifactualism original claims that in order for a literary work $w$ to exist at time $t$ there must also exist at $t$, among other things, at least one person who possesses the relevant background assumptions to understand $w$. I have shown that this condition is best understood as the claim that at least one person must be able to properly interpret $w$. Artifactualism can then be reformulated as follows: in order for a literary work $w$ to exist at $t$ there must also exist at $t$, among other things, at least one person who can properly interpret $w$.

This of courses raises many questions about what qualifies as a proper interpretation of a work and what does not. For example, if you interpret Wordsworth’s “A slumber did my spirit seal” as portraying death as a bitter, violent process while I interpret the poem as portraying death in an optimistic, pantheistic manner, has one of us misinterpreted the poem or does the poem simply admit multiple interpretations? If I read *Pride and Prejudice* and interpret it as a satire of Victorian society (as opposed to Regency society) have I misinterpreted the novel? What if I interpret *Pride and Prejudice* as taking place in the present day and satirizing strangely old-fashioned 21st century English society?

For the time being I will set these important questions aside and focus my attention on relatively uncontroversial examples of flagrantly incorrect interpretations: for example, a reader who fails to realize *Animal Farm* is a satire of Stalinism. In Chapter 5 I will take this issue up again and consider which sorts of interpretations qualify as proper interpretations of a given work and which qualify as misinterpretations.
3.3 Actual or Potential Readers?

According to my revised form of artifactualism, in order for a work \( w \) to exist at time \( t \), at least one copy of the \( w \)’s text must exist at \( t \), and at least one person who is both able to read the language \( w \)’s text is written in and who is able to properly interpret \( w \) must exist at \( t \). Before I discuss this revised form of artifactualism in more detail, there is one aspect that needs some clarification. Namely when I assert that in order for \( w \) to exist at \( t \) there must also exist at \( t \) a person who can read and properly interpret \( w \), is this person supposed to be a potential or actual reader of \( w \)? In other words, in order for \( w \) to exist at \( t \), must this reader simply possess at \( t \) the ability or potentiality to read and interpret \( w \), or must this reader actually be reading and interpreting \( w \) at \( t \)?

Thomasson favors the view that in order for \( w \) to exist at \( t \) there must exist at \( t \) at least one person with merely the potential to read \( w \). She writes:

[A] literary work requires the ongoing existence of a community capable of reading and understanding the text, it does not require that someone constantly be reading it or thinking of it in order to remain in existence, just as the ongoing existence of money requires a community willing to accept it as money although it does not constantly require that someone be explicitly thinking "this is money."\(^{326}\)

For instance, imagine that during the stretch of time beginning on April 10, 1904 and ending on June 10, 1904, not a single person in the world read *Pride and Prejudice*. *Pride and Prejudice* still existed during those two months in 1904 because countless

\(^{326}\) Thomasson, *Fiction and Metaphysics*, 23.
people who had the ability to read *Pride and Prejudice* existed during those same two months.\(^\text{327}\)

The greatest drawback to the view that only a potential reader is required to sustain a work in existence is that we lack strict criteria for when a person counts as a potential (as opposed to an actual) reader and interpreter of a work at a specific point in time. Concepts like “potential reader” and “potential interpreter” are vague, and since the existence of literary works is tightly tied to the existence of people who can potentially read and interpret them, literary works will sometimes display vagueness of existence.

For instance, it is commonly believed that the Roman emperor Claudius was the last living person who could read Etruscan.\(^\text{328}\) If a potential reader is required to sustain a work in existence, then as long as Claudius existed, literary works written in Etruscan existed. This is true even of Etruscan literary works that Claudius never actually read, since he had the potential to read them. And as soon as Claudius went out of existence and ceased to be a potential reader of Etruscan works, all Etruscan literary works went out of existence as well. But at what exact moment in time did Claudius cease to qualify as a potential reader of Etruscan works? The exact moment he died from poisoning on October 13, 54 A.D.? The moment he lost consciousness? The moment he ingested the poison? The moment years earlier when he became too busy with running the empire to ever again devote time to his previous hobby of reading Etruscan literature? Since the concept of a “potential reader” is vague, literary works sometimes have vague existence. If it is vague at \(t\) whether Claudius qualifies as a potential reader of Etruscan, then it will be vague at \(t\) whether every literary work written in Etruscan exists.

\(^{327}\) *Pride and Prejudice* also existed from April 10, 1904 until June 10, 1904 because at least one physical copy of its text existed during that time span.

Thomasson considers such vagueness inevitable but anodyne as it results from indeterminacy in ordinary language and ordinary ways of representing the world.\textsuperscript{329} However many people will not be as comfortable with literary works (or any objects) displaying vagueness of existence. If one is motivated to avoid ontic vagueness like this, one may wish to insist that in order for work $w$ to exist at $t$ there must also exist at $t$ at least one person who is actually reading and properly interpreting $w$. This move is predicated on the assumption that while it is quite vague whether or not a person like Claudius is a potential reader and interpreter or a work, there is almost no question about whether Claudius is actually reading and interpreting a work at a specific moment in time.

This move – insisting that a work exists only when it is actually being read and interpreted – comes with some tradeoffs. First by insisting that a person actually be reading and interpreting $w$ at $t$ in order for $w$ to exist at $t$, temporal gappiness in literary works becomes rampant. A work will constantly be flitting in and out of existence depending upon whether anyone is reading it at any given instant. I actually think that the increased temporal gappiness of literary works is the least troublesome consequence of insisting that an actual reader is required to sustain a work in existence. As I discuss in detail in Chapter 4, temporal gappiness in literary works really is not as metaphysically problematic as it initially appears to be. Nonetheless, many will see the increased temporal gappiness of literary works as an unacceptable drawback to requiring an actual reader of $w$ to exist at $t$.

A second drawback of taking the “actualist” position on readers of works is that it is not clear that the concepts of “actual reader” and “actual interpreter” are less vague than the concepts of “potential reader” and “potential interpreter.” There are certainly cases where it is indeterminate whether or not the predicate “actual reader and interpreter of \( w \)” would apply to a person. For example, if I observe a mentally ill man at \( t \) and think, “he reminds me of Hamlet” am I actually interpreting *Hamlet* at \( t \)?

This example points to something more fundamental about the experience of reading and interpreting literary works. The actualist position on readers requires that at \( t \) a person is both actually reading *and* interpreting \( w \). This assumes that the activities of reading and interpreting, while distinct, happen simultaneously such that a person can be said to be engaging in both activities at the same moment in time. But this is inconsonant with our normal experience of reading and interpretation. While sometimes one reads and interprets a work simultaneously, a good deal of interpretation is not simultaneous with the reading of a work. Peter Kivy points out that most literary interpretation does not happen – in fact, cannot happen – while one is reading a work. Rather the bulk of interpretation occurs during the gaps of time between reading sessions and when one reflects upon a work after one has finished reading it.\(^{330}\)

While I am not fond of the rampant ontic vagueness that flows from a “potentialist” position on readers and works, I find the actualist alternative quite undesirable. The actualist position does not entirely eliminate ontic vagueness; even worse, it requires something – simultaneous interpretation and reading – that appears to be impossible. Neither view is without problems, but since the potentialist position is less problematic, I will adopt it for the remainder of this project. The existence of a

\(^{330}\) Kivy, *Performance of Reading*, 107-114
literary work \( w \) at time \( t \) depends upon the existence at \( t \) of at least one person with the potential to read and properly interpret \( w \). I realize this will result in literary works displaying vague existence, but this is unavoidable.

3.4 Destruction of Literary Works

Literary works are very durable entities that can endure for millennia. The survival of literary works is especially impressive when compared with the survival of other products of human cultures. The grandest palaces and temples of ancient Mesopotamia and Greece - if they have survived at all - are in ruins, yet *Gilgamesh* and the *Odyssey* still exist in all their glory. From a reader’s perspective, the ancientness of literary works like the *Odyssey*, the *Psalms*, and *Beowulf*, can be awe-inducing and greatly enhances the aesthetic experience of these works. The experiences of ordinary readers and scholars with literary works, especially very old literary works, lead many to the rather romantic belief that literary works are immortal, ethereal things that, once created, are “absorbed” into the realm of eternity. This seems to be particularly true of poetry. A poem itself is immortal and also enables its subject to attain immortality; this notion is beautifully expressed in Shakespeare’s Sonnet 55

Not marble nor the gilded monuments
Of princes shall outlive this powr’ful rhyme,
But you shall shine more bright in these contents
Than unswept stone, besmeared with sluttish time.
When wasteful war shall statues overturn,
And broils root out the work of masonry,
Nor Mars his sword nor war’s quick fire shall burn
The living record of your memory.
'Gainst death and all oblivious enmity
    Shall you pace forth; your praise shall still find room
    Even in the eyes of all posterity
    That wear this world out to the ending doom.
    So, til the Judgment that yourself arise,
    You live in this, and dwell in lovers’ eyes.  

Against this background it may sound odd or jarring to argue that literary works, even Sonnet 55, can be destroyed. Thomasson notes that while folk intuition supports the artifactualist account of literary works as created not discovered entities, many people also have the intuition that once a literary work has been created it can never pass out of existence.  

She attributes this to “a hangover of a Platonism that assimilates all abstract entities to the realm of the changeless and the timeless.”  

Literary works may be abstract entities, but they are not Platonic. According to artifactualism, literary works are contingent and they can cease to exist.  

The first and most familiar way in which a literary work can become non-existent is if every single physical copy of the work becomes non-existent. This is accomplished quite easily if there is only one physical copy of the work, such as the author’s original, unpublished draft. A mere stack of papers can be burnt to ashes, a scenario played up to great dramatic effect in the film Misery. Once the protagonist’s

---

334 It must be noted that artifactualism, in principle, is not necessarily committed to the view that literary works can be destroyed. One could hold that literary works are contingent abstracta, but also hold that once a literary work has been created by its author, it exists forever. Presumably a literary work would still exist even if all copies of its text were destroyed and even if not a single person alive was able to read the language the work was written in and properly interpret the work. To the best of my knowledge no philosopher holds this view. While I fail to see the appeal of such a view, it is a logically coherent position.  
novel was burned, the novel ceased to exist because there were no remaining physical copies of it.

Ancient and medieval literary works are especially vulnerable to destruction. Before the development of the printing press and the mass production of books, physical copies of literary works were relatively rare. Only one copy of a work is required to sustain the work in existence, but one copy is more easily destroyed than one thousand copies are. Add to this the inherent difficulties of preserving objects made of parchment, papyrus, and vellum for a millennium or two, and it is no surprise that classical literary works predominate on lists of “lost works.” Among these non-existent literary works are Homer’s comic poem *Margites*, most of Sappho’s poetry, and seventy-three plays by Aeschylus.336 We know that these works existed at one point in time because other texts refer to them; Aristotle mentions *Margites*, for instance.337 In some cases we can even pinpoint the exact date at which these works became non-existent. For example, the last remaining copies of Aeschylus’ plays were burnt on orders from the Caliph of Alexandria on December 22, 640 A.D.338

Modern literary works can be destroyed as well, but this usually happens when a work has not been published yet, and there is only one physical copy of it - frequently the author’s own original manuscript. Sylvia Plath’s unpublished novel *Double Exposure* disappeared sometime during the 1960s. An untitled and unpublished war novel by Ernest Hemingway disappeared when the suitcase it was in was stolen in 1922. Gogol

himself burned his one and only copy of *Dead Souls: Part II* in a fit of religious mania at 3:00 a.m. on February 24, 1852.\(^{339}\)

Outside of extreme cases like *Dead Souls: Part II*, nowadays it is especially difficult to destroy all physical copies of a literary work, especially given the proliferation of digital information storage. In order to destroy *Pride and Prejudice*, a scenario similar to the one described in *Fahrenheit 451* would have to obtain. There are now doubtless millions of copies of *Pride and Prejudice* in bookstores, libraries, schools, and homes. Every last one of them would have to be destroyed. Every digital copy, every audio copy, and every copy in any other physical medium would have to be destroyed. Even then, it is possible some savants could memorize *Pride and Prejudice* verbatim and preserve the work in existence in their memories, just as in *Fahrenheit 451*. In order for *Pride and Prejudice* to truly cease to exist, any such savants would have to cease to exist without passing on their memory of *Pride and Prejudice* to anyone else.

It seems fairly obvious that in a situation where not a single physical copy of a literary work remains that the literary work itself has been destroyed. However, given the multiple dependencies of literary works, in some cases their destruction will not be so apparent. According to artifactualism it is possible for a literary work to cease to exist even if there are extant physical copies of it. Not only are concrete entities required to sustain a work in existence, but also specific mental states – the ability to read the work’s language and the ability to properly interpret the work – are required to sustain a work in existence. If those mental states no longer exist; that is, if no person alive has them, then the work itself will go out of existence.

\(^{339}\) Kelly, “*The Missing Masterpieces.*”
This is a somewhat counter-intuitive position. If I had to speculate, I would say that most people believe that a work of literature exists as long as there are copies of its text in existence. The notion that a work could cease to exist even if physical copies of it remained in existence does sound strange, to be sure. However if one accepts the artifactualist picture of literary works as hybrid objects, dependent upon a combination of both physical things and mental states, one cannot hold that the existence of copies of a work’s text is sufficient to sustain the work in existence. If only the loss of all copies of a work can destroy the work, and if the loss of particular mental states cannot destroy the work, then it seems reasonable to conclude that the only thing a literary work depends upon for its continued existence is at least one copy of its text. Not only does this violate the very core of artifactualism, it seems to skirt very close to the widely-discredited physical object hypothesis.

One way in which a work can be destroyed via the loss of certain mental states is if there no longer exists any person who has the language capabilities to read a literary work. In order for a literary work to exist at time $t$, at least one physical copy of the work must exist at $t$ and at least one person who can read and properly interpret the work must exist at $t$. If at $t$ there is not a single person in the world with the ability to read a work’s language, then the work does not exist at $t$, even if multiple physical copies of the work’s text exist at $t$.

Another way to put this condition is that in order for a literary work to exist at $t$, the work’s language must be a “living” language at $t$ and not a “dead” language. The term “dead language” is often used to refer to a language that is no longer used for every day speech; Latin, for example, is frequently called a dead language. But for our
purposes, I will define a dead language as a language that no person can speak or read. A living language is a language that at least one person can speak or read. So-called dead languages like Latin and Sanskrit are actually living languages under this definition. Etruscan is an example of a dead language. No person uses Etruscan for every day speech anymore, and since no scholar has been able to translate Etruscan texts, no person can read Etruscan texts. Etruscan became a dead language the moment the last person who could speak or read it died. Assuming Claudius was the last speaker of Etruscan, let us take his death in 54 A.D. to be the time at which the Etruscan language became dead.

If a literary work is written in a language that subsequently becomes a dead language, then the literary work will become non-existent the moment the language dies. “A literary work as such can exist only as long as there are some individuals who have the language capacities . . . to read and understand it.” If there are physical copies of a literary work written in a dead language these will be nothing “but some ink on paper.” The work itself no longer exists. This is so even if physical copies of the literary work remain in existence.

The demise of the Etruscan language is a real life instance of this phenomenon. Although the Etruscan language has been dead for approximately two thousand years, a fair amount of Etruscan texts still exist today. These texts usually take the form of inscriptions on stone and metal tablets. It is not unreasonable to imagine that some of these texts might be the texts of literary works - poems or plays, perhaps. Imagine one of these inscriptions is an Etruscan poem, call it “Xeqhu,” written in 300 B.C. “Xeqhu” does not exist anymore, even though a physical copy of its text exists. In order to remain

340 Thomasson, Fiction and Metaphysics, 11.
341 Thomasson, Fiction and Metaphysics, 11.
342 Thomasson, Fiction and Metaphysics, 39.
in existence, a literary work requires the presence of at least one physical copy and at least one person who can read the language the work is written in. Although “Xeqhu” has an existent physical copy of its text, there does not exist a single person who can read it because Etruscan is a dead language. Therefore “Xeqhu” no longer exists, despite appearances to the contrary. It was destroyed the moment Claudius died in 54.

3.5 Misinterpretation as Destruction

According to my revised form of artifactualism, there is one more distinctive way in which a literary work can be destroyed. A literary work depends for its continued existence on at least one copy of its text, at least one reader who can understand the work’s language and who can properly interpret the work. If there are no longer any persons with the ability to properly interpret a specific literary work, then the literary work will cease to exist. A literary work can be destroyed in this manner even if physical copies of it abound and it is written in a language easily read and understood by many people. This is a striking claim indeed. It jars our intuitions to imagine that a literary work with millions of printed copies in libraries, bookstores, and homes, further preserved through various digital media, written in a living language like English or French, and read and discussed by students, scholars, and ordinary readers does not, in fact, exist. But such is the case if not a single reader is able to properly interpret the work.

How might such a scenario come about? When imagining how such a state of affairs could obtain it seems particularly appropriate to refer to the works of George
Orwell. Let us imagine that Orwell’s nightmare has come true; at some point in the future all people are under the control of a highly oppressive totalitarian state, not unlike the one in *1984*. The state ruthlessly censors what information about history the populace is allowed to know. Among the historical figures and eras that the state is keen to censor are Stalin and the history of the Soviet Union. All written records, documents, textbooks, visual media, and digital information that refer to Stalin or the USSR have been utterly destroyed. All people with living memory of Stalin or the USSR have long since died (let us assume this Orwellian state comes into power several centuries from now). Not a single person in this future society, not even the most educated, knows anything about Stalin or the USSR. Not a single person even knows that these entities once existed. And given the thoroughness of the censorship of historical information, it is now impossible for any person in this society to ever discover facts about Stalin or the USSR.

Yet in this future dystopia, the novel *Animal Farm* is still in publication. For some reason, perhaps a bureaucratic oversight, *Animal Farm* was never censored. There are, of course, no copies of *Animal Farm* with editor’s notes or commentary that refers to Stalin or the USSR. But there are plenty of copies of the actual text Orwell wrote. English is still widely spoken in this future world, so there are millions of people who can and do read *Animal Farm* with ease.

Unfortunately not a single one of these readers is able to interpret *Animal Farm* properly. In order to interpret *Animal Farm* properly, a reader must recognize that it is a satire of Stalinism. But a reader cannot recognize a novel as a satire of Stalinism if the reader knows nothing about Stalinism. No one will be able to recognize a novel as a satire of Stalinism if nobody will ever be able to know anything about Stalinism, due to

---

the ruthless censorship and destruction of historical information. And since a literary work requires at least one reader who is able to interpret it properly, *Animal Farm* no longer exists in this situation. It ceased to exist long before, probably upon the death of the last person who had knowledge of Stalinism. Here misinterpretation has resulted in destruction - destruction of the original work.

In this future dystopia, despite appearances to the contrary, *Animal Farm* no longer exists. There are millions of physical copies of *Animal Farm*. There are thousands of English-fluent people who read *Animal Farm*, discuss it with each other, and even study it in school. But *Animal Farm* itself became non-existent a long time prior. This of course raises the question, what *does* exist in this dystopian scenario? The answer is quite simple: a text exists in the dystopian scenario, the text of *Animal Farm*. This text is not identical to the work itself, given artifactualism’s rejection of textualism. When readers and scholars attempt to interpret *Animal Farm*, strictly speaking they are not interpreting a literary work but rather a text.

3.6 An Advantage of Revised Artifactualism

For the remaining half of this chapter, I wish to discuss one of the advantages to adopting the revised artifactualist view that literary works depend upon proper interpretations to remain in existence. Up until this point, I have argued that this is a logical consequence of artifactualism. I believe one ought to be an artifactualist about literature and fictional characters, and therefore one should believe that literary works depend upon proper interpretation. However, even if one is not inclined towards
artifactualism, there is a very good reason to believe that literary works depend upon their interpretations. Revised artifactualism helps solve a puzzle about the relationship between a literary work, its text, and its interpretations.

A fine introduction to the puzzle is found in Borges’ short story, “Pierre Menard, Author of the Quixote.” The story concerns a fictional fin de siècle novelist named Pierre Menard who decides to write Don Quixote. This is an odd thing for Menard to do since Don Quixote was written in 1602 by Miguel Cervantes. Menard is aware of that but insists that he does not wish to make another copy or translation of Cervantes’ original text. Rather Menard wishes to create a new literary work entirely, his own Don Quixote. And Menard’s Don Quixote is not simply an updated re-telling of Don Quixote in a turn of the century French setting, with new characters, dialogue, and the like. Rather the text of the Don Quixote that Menard aims to and eventually does produce is word-for-word identical to the original Spanish text of Cervantes’ Don Quixote.\footnote{Technically speaking, Pierre Menard does not write a complete version of Don Quixote, but only two and a half chapters that are perfectly identical to the corresponding chapters in Cervantes’ novel. But since Borges himself treats Menard as having written Don Quixote in its entirety, and all philosophical discourse surrounding this short story treats Menard as having written an entire novel, I will treat Menard’s Don Quixote as a novel-length text as well.} It is not entirely clear from the story how Menard accomplishes this feat. Menard does not merely copy Don Quixote word for word and line by line. Nor does Menard produce Don Quixote by pretending to be Cervantes, learning Spanish, and deliberately forgetting three centuries of history. Menard does something else, left unspecified in the story, that allows him “to go on being Pierre Menard and reach the Quixote through the experiences of Pierre Menard.”\footnote{Borges, “Pierre Menard,” 40.}
How many *Don Quixotes* are there in this story, one or two? I believe the answer that one gives to this question is going to have much to do with what one takes to be the relationship between a work of literature, its text, and its interpretation.

Specifically, two competing views about the ontology of literary works – textualism and constructivism – seem to be at play in most readers’ intuitions about Menard’s *Don Quixote*. Textualism is the claim that a literary work is identical to its text; constructivism is the claim that a literary work is identical to an interpretation of it. While textualism and constructivism are mutually exclusive accounts of the ontology of literary works, both views enjoy strong intuitive support on different points. But the philosophical cases for textualism and constructivism are quite flimsy and both views are subject to serious metaphysical problems. Ultimately neither view is satisfactory.

Fortunately there is a third alternative – artifactualism, which claims that a literary work is neither identical to its text nor to any interpretation of it, but is instead dependent upon both entities in order to enjoy continued existence. Artifactualism functions as an effective via media between textualism and constructivism, incorporating the intuitive appeal of both theories while avoiding their weaknesses.

---

3.7 Textualism

Textualists claim that every work of literature is identical to its text. A text is “a sequence of symbols in a language or languages.”\(^{347}\) Not every text is a work – for example, a shopping list is a text but not a work – but every literary work has a text.\(^ {348}\) That includes works like folktales that are never written down. The sequence of words that is memorized and transmitted orally for generations functions as the text of a folktale. But typically texts are written texts, and these are the kinds of texts I will be concerned with. For instance, the text of *Pride and Prejudice* is the sequence of words that begins “It is a truth universally acknowledged” and ends some three hundred pages later with “the persons who, by bringing her into Derbyshire, had been the means of uniting them.”

It is very important to note that when I speak of a work’s text, I am not speaking of a particular copy of the work’s text, such as my personal copy of *Pride and Prejudice* or the original copy written by Austen. The text of *Pride and Prejudice* itself is something over and above individual printed copies of it. The relationship between the text of a work and individual copies of this text is probably best construed as a type-token relationship.\(^ {349}\) This abstract text type is what I am referring to when I speak of a work’s text.

\(^{347}\) Thomasson, *Fiction and Metaphysics*, 64.

\(^{348}\) However, this does not necessarily mean that every literary work is a text or that there is a neat one-to-one correspondence between works and texts.

\(^{349}\) See Kivy, *The Performance of Reading* 4; Thomasson, *Fiction and Metaphysics* 64. It is important to distinguish the claim that a work’s text is a type from the view that a literary work itself is a type. For a defense of literary works as types see Wollheim, *Art and its Objects*, 74-91; Wollheim, “Literary Works as Types,” in *Philosophy of Literature: Contemporary and Classical Readings*, eds. Eileen John and Dominic McIver Lopes (Malden, Mass.: Blackwell, 2004).
According to textualism *Pride and Prejudice* qua work is identical to the sequence of words that begins “it is a truth universally acknowledged” and ends some three hundred forty pages later with “the means of uniting them.” Under textualism there is a neat, one-to-one ratio between texts and works. Every work has one and only one text, and every text constitutes one and only one work.\(^{350}\) As a result, the textualist response to Pierre Menard’s *Don Quixote* is to simply deny that this manuscript constitutes a numerically distinct literary work from Cervantes’ original *Don Quixote*. Clearly Menard and Cervantes’ creations share the same text, call it \(q\). Before Menard created his own copy of \(q\), \(q\) was already identical to a work, *Don Quixote*. By creating another copy of \(q\), all Menard has done is create another copy of *Don Quixote* – since \(q\) just is *Don Quixote*. But that is something that can be done by anyone with access to a printing press, a photocopier, or perhaps a sufficient number of monkeys and typewriters.\(^{351}\) The reason Menard’s accomplishment is so puzzling and suggestive of two distinct works is the highly unusual and mysterious way in which Menard created his copy of *Don Quixote*. But that is all that Menard has done, create another copy of *Don Quixote*.

There are strong intuitive reasons to favor textualism. Literary works are, after all, linguistic entities. In more colloquial terms, literary works are “made out of” words, just as tables are made out of wood and wine is made out of grapes.\(^{352}\) Since a text is by

---


\(^{351}\) Goodman and Elgin, “Interpretation and Identity,” 96.

\(^{352}\) Collingwood notes that this is a common but lamentable view of literature. He finds the notion of words as the literal raw material of literature problematic because it leads to a very strange picture of authorial creation. If it is true that authors create literature from words the way smiths make horseshoes from iron then writers like Ben Johnson think something like, “I want to make a nice little hymn to open Act V, Scene VI of Cynthia’s Revels. Here is the English language, or as much of it as I know; I will use thy five
definition a sequence of words, it makes a good deal of sense on an intuitive level to view literary works as made out of or constituted by their texts. Ordinary intuition rarely questions whether constitution is identity; consider how quickly and easily folk-level beliefs identify the statue with its clay and the ship of Theseus with its planks in those and other related paradoxes. Given how ordinary intuition routinely conflates constitution and identity, it is not too far a leap – intuitively speaking – to conclude that a literary work is identical to its text.

It is also very difficult to imagine how a work could come into existence without some sort of “manifestation” as a text. How could there exist a *Pride and Prejudice* if Austen had never created the text that began “it is a truth universally acknowledged” and ended “the means of uniting them”? Such a question seems so odd precisely because intuition strongly suggests that the work *Pride and Prejudice* just is the text that starts “it is a truth” and ends “the means of uniting them.” This intuition seems to underlie the joke often shared between procrastinating students that one’s seemingly unwritten term paper or dissertation chapter is already complete, “it’s just all up here” (pointing to one’s head). The humor in this quip is due to the fact that we ordinarily do not believe that literary works (and closely related entities like dissertation chapters) can exist as mere private thoughts. Such things must be written down; such things must exist as texts in order to exist at all.\(^{353}\)

\(^{353}\) Of course our intuitions are not terribly consistent on this point. Some literary works, such as very short poems, can exist without being written down – usually in the minds of their authors or in the minds of other people who have memorized them. Ordinary intuitions also overlook the fact that in the ancient world, literary works of great length and complexity like the *Odyssey* and *Genesis* existed for centuries as unwritten oral tradition. But a memorized sequence of words still qualifies as a text, and it remains
Finally, our ordinary beliefs about the destruction of literary works seem to support textualism as well. When every surviving copy of a work’s text is destroyed, we generally believe the work itself has also been destroyed. When the censors in *Fahrenheit 451* burn all remaining copies of various novels and plays, they do not simply destroy pieces of paper with ink on them, they destroy works. We look back with sadness upon historical events like the fire that destroyed the Library of Alexandria and the ninth century Viking invasions of England that saw many a monastic library end up at the bottom of the North Sea. In addition to the human toll, these events are tragic not only because they resulted in the destruction of precious papyrus scrolls and illuminated manuscripts, but also because they resulted in the destruction of untold literary works.

There is a strong intuition, shared by bibliophiles and censors alike, that the way to destroy a literary work is to destroy all copies of its text. Such a tight connection between a work and its text seems to be more solid intuitive evidence for textualism, although in this case I think our intuitions go a step beyond textualism. Remember, textualism claims that a work is identical to its text, the text being some sort of abstract entity, a type perhaps. Textualism does not claim that a work is identical to a particular physical copy of its text or even to the collection of all physical copies of its text. But ordinary intuition seems to indicate that works must exist as texts in some physical form – usually printed books or handwritten manuscripts – in order to exist at all.

While the intuitive case for textualism is compelling, the philosophical case for textualism, as it stands today, is surprisingly weak. Consider the defense of textualism
difficult to imagine how the *Odyssey* (pre-Homer) could have existed without being made manifest as an oral text (i.e., if no person had ever spoken the words that make up the text of the *Odyssey*).

354 Although we may commonly use terminology like “lost works,” in our attitudes towards such works we treat them as formerly existent, now non-existent entities like the Colossus of Rhodes, not as merely misplaced entities like a set of keys.
put forth by Nelson Goodman and Catherine Elgin. When considering what a literary work ought to be identified with, Goodman and Elgin see only two choices: a work is identical to its text or a work is identical to an interpretation of it. If a work is identical to an interpretation of it, then every work will dissolve into multiple works, “with the interpreter at least as responsible for the work as the author is.”\textsuperscript{355} There will be dozens, perhaps hundreds of Hamlets, Ulysses-es, and Genesis-es and so on. This chaotic outcome alone would be sufficient grounds for rejecting the view that works are identical to interpretations. But as Goodman and Elgin point out, there is a further flaw. If every interpretation constitutes its own work, then it becomes impossible to claim that two competing interpretations – such as Cleanth Brooks’ pessimistic interpretation and F.W. Bateson’s pantheistic interpretation of “A slumber did my spirit seal” – concern the same work. There is no commonality between Brooks and Bateson’s interpretations of “A slumber”; these two interpretations are simply two distinct works.

Given the disastrous consequences of identifying a work with an interpretation of it, Goodman and Elgin argue that we must instead identify a work with its text. Identifying a work with its text preserves the unity of the work; there is only one “A slumber did my spirit seal” because there is only one text of “A slumber did my spirit seal.” Identifying works with their texts also accounts for the obvious fact that Brooks and Bateson’s interpretations are about the same thing. There is commonality between these two interpretations because both interpretations are directed towards the same text.

While Goodman and Elgin present good reasons to reject the claim that works are identical to interpretations, this is hardly a compelling argument for textualism. Goodman and Elgin’s argument for textualism strikes me as reactionary; textualism is to

\textsuperscript{355} Goodman and Elgin, “Interpretation and Identity,” 93.
be embraced because the alternative is so unacceptable, not because it possesses any merits on its own. As Gregory Currie points out, “textualism might seem attractive if you thought it the only straw floating. There is some evidence that Elgin and Goodman see it that way.”³⁵⁶

It is true that identifying a work with an interpretation of it has unappealing and counter-intuitive consequences; this will be discussed in more detail in the following section. But textualism also has some unappealing and counter-intuitive consequences of its own, despite its initial intuitive support. The greatest problem for textualism is that it identifies works with texts when the identity conditions of works are quite different from the identity conditions of texts.

What do I mean by this? It seems commonsensical to say that most works could undergo small changes to their texts and still remain the same work. For example, if Melville had omitted just one sentence from his chapter on whale taxonomy in *Moby Dick* or if he had named the character Starbuck “Starling” instead, it seems sensible to believe that the resulting work still would have been *Moby Dick*. But texts, being defined as a sequence of symbols in a language, cannot survive such minor changes. Currie suggests that in order for two text tokens to qualify as members of the same text type, “they should consist of words, in the same order, that mean the same thing and are spelled the same way.”³⁵⁷ The text of *Moby Dick* as it stands now and a text of *Moby Dick* that is missing a sentence about whale taxonomy or where Starbuck is named Starling are not the same text because they contain different sequences of words. If works are texts, this means that if Melville had renamed Starbuck “Starling” or had

---
decided he went on too long about narwhales and erased one sentence, this would have resulted in a completely different work. The most minor changes to a text – the omission of a word, the addition of a sentence, small changes in spelling, and the like – result in completely new literary works. This strikes me as wildly counter-intuitive.

It presents an especially odd picture of what happens when an author writes a novel. As one writes a novel, she typically makes many changes to the text – the text may begin with only one word and end as several hundred pages, certain sentences and paragraphs may be re-written many times over, here or there a line of dialogue might be changed significantly, the ending might be altered, the sequence of plot events might be re-arranged, and so on. Ordinarily we think of all this labor as directed towards just one literary work. But every minor change an author makes to her text results in a completely new and distinct work. So it is not the case that an author’s labor is directed towards one work. Rather the author constantly creates new works with every letter she writes, and the end product is the last work in a serious of perhaps millions.

3.8 Constructivism

A small family of related views in philosophy and literary theory fall under the term “constructivism,” although few of these views go by that name. Constructivism is the claim that a literary work is identical to an interpretation of it or the class of its interpretations. Perhaps a better way to put it is to say that every interpretation of a literary work constitutes or constructs a work in its own right. Brooks’ interpretation of
“A slumber” is a work, Bateson’s interpretation of “A slumber” is a work, and “A slumber” itself is nothing over and above Brooks and Bateson’s interpretations.

One notable aspect of constructivism is that it views an interpretation of a literary work as a full-fledged object in its own right. Usually interpretation is spoken of as a type of activity or event. However in the context of both constructivism and textualism, “interpretation” does not simply refer to an activity performed by readers, but also to whatever it is that is produced by such activity. When textualists and constructivists speak of a work’s interpretations, they are speaking of things that may or may not be candidates for identity with the work. So we need an account of interpretations as objects. It seems most sensible to define an interpretation of a work as first and foremost a mental state (or collection of mental states) developed by a reader of a work that aims to describe the meaning and significance of the work. Such mental states might be very complex and long-lasting – for instance, a graduate student writing a dissertation on Finnegans Wake will have a large collection of complicated mental states, developed over the course of years, that describe what she takes to be the meaning and significance of that novel.\textsuperscript{358} Other mental states might be very simple and short-lived – for instance, the high school student who reads Hamlet and briefly thinks to herself, “wow, Hamlet is nuts!”

In some cases, usually restricted to professional scholarship, an interpretation is subsequently communicated to others through writing – the graduate student’s

\textsuperscript{358} Kivy points out that a reader's interpretation of a work is something that must develop over a period of time. First of all, the reading of a work cannot take place instantaneously, it is an event that is always extended in time. And most works (except the very shortest) cannot be read in one uninterrupted sitting either. A reader’s interpretation of a work is developed over a period of time, as the reader contemplates various parts of the work in between reading sessions and after the entire work has been read. Because of this the mental states that make up a reader's interpretation of a work will almost always be temporally discontinuous. See The Performance of Reading, 107-114. See also Kivy’s “Continuous Time and Interrupted Time: Two-Timing in the Temporal Arts,” Musicae Scientiae Forum 3 (2004), 141-155.
dissertation, for example. And many times an interpretation is subsequently communicated to others orally; the high schooler might mention her insight to her friend during lunch. But most of the time, interpretations remain in the private realm of readers’ mental states.

Constructivist intuitions appear to be at play among people who read Borges’ story and conclude that Menard has, in fact, created a new and distinct work. Why is Menard’s *Don Quixote* different from Cervantes’ *Don Quixote*? The answer lies in the fact that two different interpretations attach to each text. Borges’ narrator holds this view and defends Menard’s accomplishment against skeptics by pointing out that Menard’s *Don Quixote* is ironic, ambiguous, and influenced by William James, whereas Cervantes’ *Don Quixote* is unironic, unambiguous, and obviously not influenced by James.359 But to claim that *Don Quixote* is ironic or influenced by James is to say something about the meaning and significance of that work - it is to interpret *Don Quixote*. Different interpretations – such as “*Don Quixote* is ironic” vs. “*Don Quixote* is unironic” – result in different works, despite the fact that these works share one text. A work’s identity does not lie with its text but rather with its interpretation.

It is important to distinguish between radical and moderate constructivism. Radical constructivism claims that every individual interpretation is identical to a work, while moderate constructivism claims that a work is identical to the class of its interpretations. Radical constructivism does not appear to be advocated by any philosopher, but rather has its genesis in recent critical theory. Roland Barthes seems to hint at radical constructivism in “The Death of the Author” with his claim that every literary work is “eternally written *here and now*” as each individual reader reads and

gives meaning to its text. More explicit advocacy of radical constructivism is found in
the reader-response criticism of Stanley Fish. Fish rejects the commonsense view that
critical attempts to interpret a work, such as Milton’s Lycidas, are all directed towards
some thing – “Lycidas itself” – that exists independently of the readers who interpret it.
Rather “it is the structure of the reader’s experience rather than any structures available
on the page” that interpretation is actually directed towards. The “work itself” is the
reader’s experience of it, the reader’s interpretation of it. When one interprets a work,
one literally creates a work of one’s own. According to Fish, when he reads Lycidas and
interprets it as a defense of fantasy, this enables him to “write the text I write when
reading Lycidas” If another reader were to interpret Lycidas differently from Fish,
then “we could not possibly be reading the same poem . . . for each of us would be
reading the poem he had made.”

Moderate constructivism’s most notable advocate is Michael Krausz. Krausz’s
moderate constructivism is motivated by his belief that literary interpretation is often
“imputational.” This means that interpretations can add properties to literary works; in so
doing, interpretations help to constitute or construct literary works. Since many
different and competing interpretations can impute properties to a literary work, literary
works “are the class of their interpretations . . . interpretation constitutes objects of

Irwin (Westport, Conn.: Greenwood Press, 2002), 5.
361 Stanley Fish, “Interpreting the Variorum,” in Is There a Text in this Class? The Authority of Interpretive
362 Fish, “Interpreting the Variorum,” 159.
363 Fish 169, “Interpreting the Variorum,” emphasis in original.
364 Fish 169, “Interpreting the Variorum,” emphasis added.
interpretations.” By object of interpretation Krausz means a literary work, although his theory of interpretation also applies to other works of art like paintings and music.

Constructivism lacks the broad intuitive support that textualism enjoys; more people read *Fahrenheit 451* than “Interpreting the Variorum” after all. But among professional scholars of literature and critical theorists, constructivist-friendly intuitions are not uncommon. Chief among these intuitions is the view that works of literature are somehow inert or incomplete unless they are read and interpreted. Georges Poulet describes the exciting metamorphosis that occurs when he picks up a book and begins to read and interpret it:

Where is the book I held in my hands? . . . The book is no longer a material reality. It has become a series of words, of images, of ideas, which in their turn begin to exist. And where is this new existence? Surely not in the paper object. Nor surely, in external space. There is only one place left for this new existence: my innermost self.

By reading and interpreting a book, Poulet gives the literary work “existence.” This sentiment is echoed by Hans Jauss who finds the “life of a literary work . . . unthinkable without the active participation of its addressees” and Wolfgang Iser who argues that “the convergence of text and reader brings the literary work into existence.”

It is difficult to speculate how common these intuitions are outside of English departments. But even if these intuitions are rare at the folk level it is still appropriate to consider the intuitions of people who interpret literature for a living when trying to work

---

368 Poulet, “Phenomenology of Reading,” 59.
out the ontology of literary interpretation. Many professional critics are of the opinion that interpretation plays an extremely important role in the very existence of a literary work. If a work like *Hamlet* were never interpreted, then *Hamlet* qua work would either not exist at all or would exist in some attenuated and unrecognizable form. From this point of view, constructivism makes a good deal of sense. The reason why *Hamlet* would fail to exist if no one interpreted it is because *Hamlet* simply is either an interpretation of it or the class of its interpretations.

Despite some intuitions in its favor, constructivism – like textualism – is subject to many counter-intuitive consequences. As Goodman and Elgin point out, if radical constructivism is true then there is not one *Hamlet* or one *Ulysses* but rather hundreds, perhaps millions of *Hamlets* and *Ulysses*-es. Every reader and every critic produces an interpretation and if every interpretation constitutes a work, then for every reader there will be a numerically distinct work. To make matters worse, if a work is nothing more than an individual creation – i.e., *my* *Hamlet*, *your* *Hamlet*, Mary’s *Hamlet* – then it can never be the case that two people read the same work. To top it all off, neither Mary, nor you, nor I ever read Shakespeare’s *Hamlet*.

The radical constructivist claim that by interpreting a work one is literally creating – writing – a work of one’s own strikes me as wildly implausible. Consider the difference between the creative activity of an author, such as Austen, and the “creative

---

371 I equivocate about the number of *Hamlets* because radical constructivism is unclear as to whether there is a literary work correlated with each interpretation type or each interpretation token. That is, if Mary and Mike both interpret *Hamlet* such that Hamlet is insane, while Joe interprets *Hamlet* such that Hamlet is not insane, it is unclear if radical constructivists like Fish would count two *Hamlets* (for two interpretation types) or three *Hamlets* (for three interpretation tokens). Since radical constructivism is a view that comes to us from literary theory and not philosophy, it is unsurprising that this matter has been overlooked and left undetermined. If interpretation tokens are works then there will be millions of *Hamlets*, where as if interpretation types are works then the number of *Hamlets* will be more modest, probably in the low hundreds. Either way, however, there are going to be far more *Hamlets* than the number – one – that common sense and professional scholarship allows.
activity” of a reader, Mary, who interprets *Pride and Prejudice*. When Austen created *Pride and Prejudice*, she composed a text, brought new fictional characters into existence, developed an original storyline, and expressed her views about English society in a unique and satirical manner. The creation of *Pride and Prejudice* was the result of immense creativity, originality, talent with language, and undoubtedly intensive intellectual labor on the part of Jane Austen. When Mary reads and interprets *Pride and Prejudice*, she has certainly created some thing, her interpretation of *Pride and Prejudice*. She has brought into existence a mental state or set of mental states, perhaps complex, perhaps very simple, that more or less describe what Mary takes to be the meaning and significance of *Pride and Prejudice*. But this act of creation on Mary's part displays few of the features and few of the demands of Austen’s. In no way can what Mary has done be considered the creation, let alone the writing, of a literary work.372

Moderate constructivism also faces serious problems. In particular, moderate constructivism leaves us unable to explain why Brooks and Bateson’s interpretations of “A slumber did my spirit seal” are interpretations of “A slumber” and not of, say, “Ode to a Grecian Urn.” Normally we would say that Brooks and Bateson’s interpretations belong to the same class of interpretations because they are both interpretations of something independent of that class, “A slumber” qua work. But under moderate constructivism, such an explanation is incoherent because there is no work “A slumber”

---

372 Translations provide an interesting exception to this rule. In the case of translation, the translator's interpretation of the original work is accompanied by the creation of a new work. For instance when Seamus Heaney translated *Beowulf*, he had to interpret *Beowulf*. He had to make decisions about what English words best expressed the connotations of the original Old English words. He had to make decisions about when to sacrifice the literal meaning of a word in order to preserve the poem's rhythm and vice versa. Guiding all such decisions was Heaney's view about the meaning and significance of *Beowulf* (or of specific parts of *Beowulf*). In other words, Heaney interpreted *Beowulf*. But in this case Heaney's interpretation of *Beowulf* constitutes a work in its own right. Heaney composed a text, and he composed it with the intention to create an English translation of *Beowulf*. 
independent of the class of its interpretations. It remains unanswered why Brooks and Bateson’s interpretations constitute “A slumber” and no other work.  

3.9 A Via Media

We are at an impasse. Both textualism and constructivism enjoy strong intuitive support. On one hand many people believe that literary works are “made out of words” and that if every copy of a work’s text is destroyed then the work itself is destroyed too – intuitions that suggest textualism. On the other hand, some people – particularly professional critics – believe that literary works cannot exist or cannot exist in their “fullest” form unless some person reads and interprets them – intuitions that suggest constructivism. But the philosophical cases for textualism and constructivism are not strong and both views lead to unacceptably counter-intuitive consequences.

Perhaps the problem with both textualism and constructivism is that they are overly reductionist regarding literary works. Textualists wish to reduce the identity of a work to its text and constructivists wish to reduce the identity of a work to its interpretations(s). But as Currie asks, “why insist on a reductive identification of works? Why not simply say that there are works?” I think Currie is correct to claim that literary works exist and cannot be reduced to texts or interpretations, but to leave the matter at that is philosophically unsatisfying. Intuition strongly suggests that both a work’s text and a work’s interpretations are extremely important for the very existence of that work. Simply saying that there are literary works that reduce to nothing ignores

\[373\] Wollheim uses a similar line of argument to show why art works like novels and operas cannot be identified with the class of their copies or performances. See Art and its Objects, 8-9.

these intuitions, which surely must have *some* relevance to the ontology of literary works. Can we provide a philosophically richer account of literary works that incorporates both textualist and constructivist intuitions while avoiding the pitfalls of both theories?

Here is my suggestion: a literary work cannot be reduced to its text, nor can it be reduced to any one (or the class) of its interpretations. But once a literary work has been created by its author, it depends upon the existence of at least one copy of its text and the existence of at least one person who can properly interpret that text in order to enjoy continued existence. This is the account of literary works put forth by my revised form of artifactualism.

Artifactualism claims that literary works exist and reduce neither to texts nor to interpretations. However in order for a literary work to exist once it has been created by its author, there must exist at least one copy of the work’s text and at least one person who can properly interpret the text. A work is not identical to its text nor to its interpretations but rather is *dependent* upon both. With this account of the relationships between works, texts, and interpretations, artifactualism functions as an effective via media between textualism and constructivism, respecting the intuitions friendly towards both theories but avoiding the unacceptable consequences of both theories too.

The philosophical case for textualism is flimsy but there are strong intuitions in favor of the view. For instance we have a hard time conceiving of how a work could exist without being manifested as a text; we laugh off the suggestion that a dissertation chapter could exist solely in one’s mind. From an artifactualist perspective this intuition is absolutely correct; a literary work cannot exist unless there is a physical copy of its text in existence. However this is not because works are their texts, but rather because works
depend upon their texts for their existence. Likewise the intuition that the destruction of every copy of a work’s text destroys the work is also correct, but correct because a work depends upon the existence of at least one physical copy of its text, not because the work is identical to the copy or class of copies. Where textualists postulate identity, artifactualists see dependence.

Textualism’s most serious flaw—minor changes to a text result in completely separate works—is easily dodged by artifactualism. Unlike textualists, artifactualists do not claim a strict one-to-one ratio between works and texts. Instead a work can be sustained in existence by any one of a number of slightly different texts. *Moby Dick’s* continued existence can be sustained by a text that contains the name “Starbuck,” or a text containing “Starling.” These two distinct texts count as texts of *Moby Dick* just as two distinct printed books—one on my bookshelf and one in the library—count as copies of *Moby Dick*. Obviously there are going to be some limits on how different two texts can be before they are no longer texts of the same work; no artifactalist would want to claim, for instance, that the text of *Pride and Prejudice* could count as a text of *Moby Dick* and sustain the latter work in existence. The best way to draw the boundaries between texts that count as copies of a work and texts that do not is to require the former class of texts to be connected to the author’s original manuscript by some chain of copying. If a publisher’s error results in a text containing “Starling” instead of Starbuck then this text would still count as a text of *Moby Dick* because the publisher

375 Thomasson, *Fiction and Metaphysics*, 64-65. Of course there will be some limits on these chains of copying. For example, imagine a chain of copying that functions like an extreme case of the game “telephone.” The chain of copying starts with Melville’s original manuscript of *Moby Dick* and replaces one word with a new one each time the text is copied. After millions of iterations, a text identical to that of *Pride and Prejudice* is produced. We would not want to say that this text counts as a copy of the text of *Moby Dick*, despite the chain of copying connecting it to Melville’s original manuscript.
intended to make another copy of Melville’s original manuscript. But if the same publisher produces a text that contains all the words commonly found in *Pride and Prejudice* this text would not count as a text of *Moby Dick* because there would exist no chain of copying between this text and Melville’s original manuscript. Ultimately works are individuated by their authors in artifactualism, not by their texts.

As for constructivism, artifactualism also respects its underpinning intuitions. Poulet and other critical theorists are correct to see unread and uninterrupted literary works as somewhat non-existent or incomplete. This is not because literary works are their interpretations, but rather because literary works depend upon the existence of at least one proper interpretation – i.e., the existence of at least one person who has some mental state that counts as a proper interpretation of the work – in order to exist. If a text of a literary work, such as *Animal Farm* in the hypothetical scenario, can only be misinterpreted then the work no longer exists, even if many printed copies of its text do. Poulet and others are correct to see works as more than just ink marks, paper, and binding. Works do require the existence of people, of readers, of interpreters, in order to remain in existence.

But where constructivists see identity, artifactualists postulate dependence. This enables artifactualism to avoid the many problems that befall both radical and moderate constructivism. Radical constructivism entails that literary works like *Hamlet* dissolve into hundreds, perhaps millions, of *Hamlets* due to the fact that every individual reader’s interpretation constitutes a literary work. Artifactualism obviously does not face this problem; for artifactualists there is only one *Hamlet* which depends upon there being at least one correct interpretation, but is never identical to that interpretation. Artifactualism
also avoids radical constructivism’s bizarre conception of ordinary reading and interpreting as equivalent to writing a literary work. According to artifactualism, a work is created when an author or authors compose a text with the intention of creating a literary work. An ordinary reader or scholar who reads and interprets a work obviously does not compose a text, nor does he intend to create a literary work, despite Fish's claims.

Artifactualism is also spared the metaphysical problems that arise from moderate constructivism. The main problem faced by Krausz’s version of constructivism – there is no way to explain why two different interpretations are interpretations of the same work – is due to Krausz’s identification of a work with the class of its interpretations. But artifactualism does not reduce the identity of a work to the class of its correct interpretations. Rather artifactualism requires that the class of a work’s correct interpretations have at least one member at time \( t \) in order for the work to exist at \( t \). But once again, dependence is not identity.

By rejecting reductive theories of a work's identity in favor of a detailed account of a work's dependencies, artifactualism functions as an effective via media between textualism and constructivism. A work is not identical to its text, nor is it identical to its interpretations. Rather every work depends upon there being at least one copy of its text and at least one proper interpretation of it in order to exist at any given moment.
3.10 Concluding Remarks

While Thomasson does not go into detail about what is meant by the proper “background assumptions” a reader must have to understand a work, I have shown that this condition is best understood as the claim that a reader must be able to properly interpret a work. Under my revised form of artifactualism, a literary work depends for its continued existence upon there being at least one copy of its text and at least one person with the potential to read and properly interpret it. As a result, in certain situations the misinterpretation of a literary work will result in its destruction, even if physical copies of its text still exist.

My claims about artifactualism and interpretation are a logical outgrowth from artifactualism itself. I believe there are solid reasons to adopt artifactualism, but even if one is not inclined to fully embrace artifactualism, there are still advantages to be found in the revised artifactualist account of interpretation. The claim that a literary work is dually dependent upon both a copy of its text and its proper interpretation offers a novel solution to the debate between textualism and constructivism. By explaining a work’s relationship to its text and interpretations in terms of dependence and not identity, artifactualism is able to respect the intuitions that make both textualism and constructivism initially so appealing, but avoid the unappealingly counter-intuitive consequences of both reductive views.
The previous three chapters have been devoted to defending the artifactualist account of fictional characters and literary works. In this chapter I will assume that artifactualism is true and focus on a very specific aspect of it. Perhaps some readers have already noted that given artifactualism’s claim that literary works depend for their continued existence upon a combination of spatio-temporal entities and mental states, artifactualism entails that literary works will be temporally gappy entities. Here I will discuss in detail why artifactualism results in temporal gappiness and why this gappiness is so troublesome for artifactualists. I will then examine some strategies artifactualists might use to avoid gappiness in literary works. As we will see, these strategies are ultimately not successful. However this should not cause despair; it turns out that gappiness is not as problematic for artifactualists as previously believed.

4.1 Temporal Gappiness

First things first; the notion of a *temporally gappy* object must be defined. An object $x$ is temporally gappy (or simply is gappy since I am not concerned with spatial gappiness) if and only if $x$ exists during some interval of time $i_1$, $x$ does not exist during some subsequent interval of time $i_2$, and $x$ exists during some further subsequent interval
of time $i3$. The $x$ that exists during $i1$ is identical to the $x$ that exists during $i3$. The intervals of time can be instantaneous or they can be centuries long; it does not matter.

In more colloquial terms, temporally gappy objects are often described as things that “flit in and out of” or “pop back into” existence. The connotations of “flit,” “pop,” and other words frequently used to describe gappy objects indicate the poor reputation gappy objects have long had in the eyes of metaphysics. One of the earliest, and probably most well-known, critiques of temporal gappiness is found in the standard objection to George Berkeley’s idealism. If Berkeley is correct that ordinary concrete objects are really nothing more than bundles of ideas and perceptions, then objects will cease to exist as soon as no one perceives them and immediately come back into existence as soon as someone perceives them. Berkeley himself realized that temporal gappiness in concrete objects was a serious drawback to his view. Keen to avoid saying that a table, for instance, constantly goes in and out of existence, Berkeley asserted that God always perceives the table. The table exists at three in the morning when no person is in the study to perceive it because God perceives it.\(^{376}\) This move led to the famous limerick by Ronald Knox:

There was a young man who said God,
   must find it exceedingly odd
   when he finds that the tree
   continues to be
   when no one's about in the Quad.

Dear Sir, your astonishment's odd
I'm always about in the Quad
And that's why the tree

continues to be
Since observed by, yours faithfully, God.\textsuperscript{377}

Berkeley’s \textit{deus ex machina} maneuver is part of a long philosophical tradition of looking askance at temporally gappy objects. And while gappiness has typically been seen as quite bad when manifested by tables and trees, temporal gappiness in persons is viewed as even more problematic. To give just one famous case, avoiding temporal gappiness in persons was a major motivating factor for Locke to adopt memory – rather than consciousness or a Cartesian soul – as the foundation of personal identity.\textsuperscript{378}

For the most part, gappiness’ poor reputation is warranted. I strongly believe that persons cannot be temporally gappy, and I am almost as certain that concrete entities cannot be gappy either. However this does not necessarily mean that literary works and fictional characters cannot be temporally gappy. Literary works are not concrete objects, nor are they identical to authorial mental states; they belong to an entirely separate category, abstract artifacts. The fact that Jane Austen cannot be temporally gappy and that the copy of \textit{Pride and Prejudice} sitting on my bookshelf cannot be temporally gappy, does not entail that \textit{Pride and Prejudice} itself cannot be temporally gappy.

4.2 Gappiness and Literary Works

Closer consideration of artifactualism’s account of literary works suggests that there are many circumstances under which a literary work can be gappy. This is due to the fact that, according to artifactualism, literary works depend upon a combination of

concrete objects and intentional states in order to enjoy continued existence. A literary work exists at time \( t \) if and only if at \( t \) at least one physical copy of its text exists and at least one person who can read the text’s language and properly interpret the text exists. If any one of these three conditions is absent, then the literary work does not exist at \( t \).\(^{379}\)

But these conditions – the presence of at least one copy of the text and the presence of at least one qualified reader – appear to be the kinds of conditions that can easily obtain during one interval of time, not obtain during a subsequent interval of time, and obtain again at some later interval of time. If that occurs then the literary work will be gappy, for it will exist during one interval of time, not exist during a subsequent interval of time, and exist again during a second subsequent interval of time.

One way in which a literary work might be gappy is if the sole copy of the work’s text is apparently irretrievably lost only to be rediscovered many years after. For instance, I have already spoken of some literary works, such as Aeschylus’ seventy-three plays as non-existent. In fact I have even specified the exact day on which these seventy-three objects passed from existence into non-existence. But, the objection continues, perhaps the historical account of what happened on that fateful day in Alexandria is incorrect on one important point. The copy of Aeschylus’ works that the Caliph ordered burnt was not the last remaining copy of those plays. What if there was another copy of the seventy-three plays and this copy still exists? What would happen if this copy were to be discovered next week, perhaps in the so-called Villa of the Papyri near Pompeii?

This villa contained an extensive library of papyrus scrolls and was buried during the 79 A.D. eruption of Mt. Vesuvius. The scrolls survived the eruption, albeit in a charred state, and with the aid of CT scanning can finally now be read again. As much of the

library is unexcavated, many more scrolls await discovery. What if Aeschylus’ lost plays were among those scrolls? A scroll with one of Aeschylus’ plays on it would count as a copy of the play. Would the discovery of this scroll mean that the play that was destroyed in 640 A.D. exists again? If so, the objection goes, Aeschylus’ plays will be gappy objects. The seventy-three plays came into existence in the 5th century B.C. when Aeschylus composed them. They went out of existence in 640 A.D. Then in 2011 the plays came back into existence when another copy was found.

Such a scenario is plausible – ancient texts believed to have been lost forever are occasionally unearthed – but it does not indicate that such a literary works are temporally gappy. A distinction needs to be made between lost works and destroyed works. Ordinary language refers to works like Margites and the seventy-three plays as “lost” but philosophically speaking this is not accurate. To call a work a “lost” work implies that it still exists but it has been misplaced and cannot be perceived without difficulty, just as a “lost” set of keys still exists but cannot be perceived without difficulty. As far as we know, Margites and the plays of Aeschylus are not lost, they have been destroyed. As far as we know, these literary works do not exist anymore anywhere at all. But this is only as far as we know. It is certainly possible that a copy of the seventy-three plays may turn up next week in the Villa of the Papyri or somewhere else. If that were to happen, then those plays will exist next week, but those plays also exist today, and have existed since the fifth century B.C. These plays are lost, not nonexistent.

Contrary to what I said earlier, the seventy-three plays were not destroyed on December 22, 640 A.D. Instead a copy of the seventy-three plays was destroyed that

---

day, but this did not result in the destruction of the plays themselves. Another copy of
the plays, the one in the Villa of the Papyri, remained in existence. The fact that no
person could perceive this physical copy from 79 A.D. until 2011 matters not, for
physical objects can persist through time without being perceived. Aeschylus’ plays are
not gappy objects, passing out of existence in 79 A.D. or 640 A.D. and popping back into
existence in 2011. These plays came into existence in the 5th century B.C and have
remained in existence up until the present moment.

While this sort of scenario does not indicate that literary works are gappy, it is an
important reminder to artifactualists to be careful about how we use language. Unless we
have compelling historical evidence that every single copy of a literary work has been
destroyed (such as the case of Gogol’s Dead Souls: Part II), we should speak of
“apparently destroyed” or “apparently non-existent” works instead of “destroyed” or
“non-existent” works. Margites and Aeschylus’ seventy-three plays have apparently
been destroyed, but there is always room for optimism. As Thomasson notes,

Who knows what may be lurking in the basement corridors of the Bodleian
Library? A formerly lost sonnet of Shakespeare’s was discovered there not long
ago. Unlike in the case of a unique painting, of which we can find the ashes, we
can always hold out hope in the case of a literary work that a copy of it remains in
some library, attic room, or perfect memory, so that the literary work might be
“found” again. \(^{381}\)

By being more cautious about how we use terms like “lost” and “nonexistent,” it
is relatively easy to avoid temporal gappiness in literary works when their texts are lost
and subsequently rediscovered. However it is much more difficult to avoid construing

\(^{381}\) Thomasson, Fiction and Metaphysics, 10.
literary works as temporally gappy when the other entities that literary works depend upon – mental states – go in and out of existence.

According to artifactualism a literary work exists only as long as at least one person who can read the language the work is written in exists. Another way of stating this is that a literary work exists only as long as the work’s language exists. Literary works can become nonexistent via the deaths of their languages even if multiple physical copies of the works’ texts remain in existence.

Human languages are quite prone to temporal gappiness. Probably the most famous example of a gappy language, one that nearly every high school student learns about, is ancient Egyptian and its hieroglyphic script. Ancient Egyptian was a living language for several thousand years and many hieroglyphic texts from that period survive to this day. But by 400 AD or so, ancient Egyptian had become a dead language. No one alive spoke it, and no one knew how to read hieroglyphics. Over the centuries a few attempts were made to decipher hieroglyphics, but no scholar had any success until 1799 when Napoleon’s troops discovered the Rosetta Stone.\textsuperscript{382}

The Rosetta Stone is a large stone dating back to 196 BC. It is inscribed with an ancient Egyptian hieroglyphic text that discusses a tax repeal. Underneath the hieroglyphics is a Greek translation of the Egyptian text. At some point after its creation, the Rosetta Stone became buried under a wall. Since the Rosetta Stone acted as a Greek-Egyptian dictionary of sorts, and since Greek was a living language in 1799, scholars were finally able to translate ancient Egyptian hieroglyphics. The actual decipherment of hieroglyphics did not occur until the work of Jean-Francois Champollion in the 1820s.\textsuperscript{383}

Since I will refer to Champollion’s work frequently throughout this chapter, for simplicity’s sake I will take the year 1824 – the year Champollion published his findings as *Precis du systeme hieroglyphique*\(^{384}\) – as the time when ancient Egyptian became a living language. Ancient Egyptian became a living language again in 1824 because in 1824 at least one person could read ancient Egyptian. Another way of stating this is that ancient Egyptian came to exist again in 1824 due to the work of Champollion.

However this means that ancient Egyptian language is a temporally gappy entity. It came into existence some time before 2000 BC and went out of existence around 400 AD. From 400 AD until 1824, ancient Egyptian did not exist. After 1824 it began to exist again.

Since any literary work depends upon the existence of its language, if a language is gappy then literary works written in that language will be gappy as well. Ancient Egyptian is a gappy language, so all ancient Egyptian literary works are gappy entities. Take for instance the short story *The Tale of Sinuhe*. This was composed around 1960 B.C., and a papyrus containing this work’s text has survived to the present day.\(^{385}\) Even though there has always been a physical copy of *The Tale of Sinuhe* in existence, the literary work itself ceased to exist in 400 A.D. when ancient Egyptian became a dead language. But in 1824 when ancient Egyptian became a living language again, *The Tale of Sinuhe* came to exist again. Thus *The Tale of Sinuhe* existed for an interval of time (1960 B.C. - 400 A.D.) did not exist for a subsequent interval of time (400 A.D. - 1824), and then existed again during yet another interval of time (1824 to the present moment). That is, of course, the definition of a gappy object.

As I argued in Chapter 3, artifactualism is also committed to the view that a literary work’s continued existence depends upon the existence of at least one reader who can properly interpret the work. The dependency of literary works upon readers who can properly interpret them can also lead to literary works being gappy. The way in which this might occur closely parallels literary works that are gappy due to the gappiness of their languages, such as *The Tale of Sinuhe*.

Consider the “dystopian scenario” from Chapter 3 in which at some point in the future, say 2600 A.D., *Animal Farm* no longer exists. *Animal Farm* does not exist in this scenario because there is not one single person capable of reading its text and interpreting it properly. All information relating to the history of the Soviet Union has been destroyed by the government, leading to no one being able to recognize that *Animal Farm* is a satire of Stalinism. Printed copies of the novel still exist and plenty of people who can read English still exist. But everyone who reads *Animal Farm* believes it is simply a strange story about barnyard animals and nothing more. This is a gross misinterpretation of *Animal Farm* and since no one is capable of interpreting *Animal Farm* correctly in 2600, *Animal Farm* no longer exists in 2600. But now let’s imagine that a history book dating from 2011 and containing a detailed account of the history of the Soviet Union escaped the initial rounds of government censorship. Perhaps it was forgotten about in someone’s attic. It lies undiscovered and unread for several hundred years. But in 2700 the book is rediscovered. Within five years literary scholars are able to put two and two together and realize that *Animal Farm* is not a strange story about animals but rather a satire of specific historical events and persons. In 2705 there exists at least one person who can properly interpret *Animal Farm*. Thus *Animal Farm* existed for an interval of time (1948
- 2600), did not exist for a subsequent interval of time (2600 - 2705) and then existed again during yet another interval of time (2705 and onward). Once again, this is the definition of a gappy object.

So it appears that if one accepts the artifactualist account of literary works, one will have to accept that sometimes a literary work will be temporally gappy. While it is relatively simple to avoid gappiness in literary works when a physical copy of a text is lost and rediscovered, it is quite difficult to avoid it in situations where a language or a proper interpretation of a work is lost and found again. Given the poor reputation gappy objects have among metaphysicians, should artifactualists be concerned about this?

4.3 Why Gappiness is a Problem for Artifactualism

For her part, Thomasson is not concerned about temporally gappy literary works. In the case of literary works written in a language that is lost and then recovered via a Rosetta Stone-style discovery, gappiness is an unavoidable but not particularly troubling consequence. She writes:

The possibility always remains, however, that these stories, and the characters represented in them, may be brought back on the basis of these texts if the language is once again discovered and understood. That one and the same character or story may exist, cease to exist, and exist once more is not so strange. Given that fictional characters and stories are not spatiotemporally located entities, there seems no reason to require spatiotemporal continuity as an identity condition.\(^{386}\)

\(^{386}\) Thomasson, *Fiction and Metaphysics*, 154-155 n. 7.
Nonetheless, I believe Thomasson is mistaken to dismiss the troublesome consequences of gappiness so quickly. First of all, it is surprising that she describes literary works (stories) as “not spatiotemporally located entities.” Literary works are, of course, not spatial - they are abstract, not concrete - but they certainly are temporally located. A literary work comes into existence at a specific point in time and if it goes out of existence, it does so at a specific point in time. The Tale of Sinuhe for instance, was created in 1960 B.C. and existed until it was destroyed in 400 A.D. It came into existence again in 1824. Since creation is simply an object’s passing from a state of non-existence into existence, we can also say The Tale of Sinuhe was created in 1824 since it passed from non-existence into existence in 1824. And presumably The Tale of Sinuhe that came into existence in 1960 B.C. is identical to The Tale of Sinuhe that came into existence in 1824.

Perhaps the problem with this scenario is already apparent. Recall that Thomasson believes that literary works have their origins essentially. Not only is being created by James Joyce an essential property of Ulysses, but also being created by James Joyce in 1922 is an essential property of Ulysses as well. Had Joyce composed the same text in 1916, this text would be a numerically distinct work from Ulysses. Likewise, being created by its author in 1960 B.C. is an essential property of The Tale of Sinuhe. Any story that lacks this property will not be identical to The Tale of Sinuhe. But if The Tale of Sinuhe comes back into existence – that is, is created in 1824 – then whatever story comes came into existence in 1824 will not be identical to the story created in 1960 B.C. These two stories have different origins and since origin is an essential property of a literary work, these two stories are two distinct literary works. There is no gappy object

387 Thomasson, Fiction and Metaphysics, 8-9.
here; given artifactualism’s insistence upon the essentiality of a work’s origins, it is impossible for literary works to be gappy.

If it is impossible for literary works to be gappy, what are artifactualists to make of cases like *The Tale of Sinuhe*? I believe artifactualists have but two options in these sorts of cases. The first option is argue that the reason why *The Tale of Sinuhe* is not gappy is because there was only one interval of time during which it existed, 1960 B.C. until 400 A.D. After 400 A.D., *The Tale of Sinuhe* no longer existed and this was a permanent state of affairs. Whatever story came into existence in 1824 when Champollion translated the hieroglyphic text of *The Tale of Sinuhe* was not identical to *The Tale of Sinuhe* created in 1960 B.C. The second option is to argue that the reason why *The Tale of Sinuhe* is not gappy is because there was no interval of time since 1960 B.C. during which it did not exist. *The Tale of Sinuhe* was only apparently nonexistent from 400 to 1824; in actuality it has existed without interruption since its creation in 1960 B.C. up until the present moment.

The first option strikes me as quite implausible. It is difficult to imagine why, given artifactualism’s conditions for the continued existence of literary works, a work like *The Tale of Sinuhe* would not exist in 1824. All that is required for a literary work to exist once it has been created by its author is the presence of at least one physical copy of its text and the presence of at least one person who can read the text’s language and properly interpret the text. All of these conditions were present in 1824. At least one physical copy of *The Tale of Sinuhe* existed, a well-preserved papyrus scroll. And at least one qualified reader existed in 1824 – Champollion. Champollion could read the language *The Tale of Sinuhe* was written in and presumably Champollion was able to
interpret the story properly. It would be incoherent for an artifactualist to claim *The Tale of Sinuhe* did not exist in 1824.

The second option is appears more promising and is worth exploring in detail. An artifactualist who embraces this option needs to explain how, despite appearances to the contrary, literary works like *The Tale of Sinuhe* have existed uninterruptedly since their authors created them. Specifically the artifactualist who follows the second option must give an account of how *The Tale of Sinuhe* managed to remain in existence from 400 until 1824 while there was not one single person who was able to read ancient Egyptian hieroglyphics during that time period.

4.4 The Anti-Gappiness Strategy

The *anti-gappiness strategy* is the name I am giving to what I think is the strongest line of reasoning an artifactualist who wishes to avoid gappiness at all costs can adopt. This strategy involves explaining how a literary work can continue to exist without interruption even when the language the work is written in has, by all appearances, become dead or non-existent. The key to the anti-gappiness strategy is a very liberal account of what it means for a person to be a *potential reader* of a literary work. As I discussed in Chapter 3, the concept of a potential reader is quite vague and under-defined. However, I see no reason why this should be seen as a prohibition against any philosophical attempts to provide a more detailed account of “potential reader.”

The distinction between actual readers and potential readers of a literary work is an important one in artifactualism. According to Thomasson, a literary work requires the
presence of at least one person who can read it to remain in existence. In order for *Ulysses* to exist at *t*, at least one person who can read and understand *Ulysses* must exist at *t*. This person need not be anyone in particular; I can be the potential reader, you can be the potential reader, Amie Thomasson can be the potential reader, and so on. As long as there is at least one person who can read *Ulysses*, it does not matter who that person is. But note that this is a person who *can* read *Pride and Prejudice*; this person is a potential reader. In order for *Ulysses* to exist at *t*, it is not necessary for this person to actually be reading *Ulysses* at *t*; this person simply has to have, at *t*, the ability to read *Ulysses*.

The view that a potential, not actual, reader is required to sustain a work in existence is not without its problems. As I mentioned in the previous chapter, “potential reader” is a vague concept that leads to ontic vagueness. But the only alternative is to require an actual reader and interpreter to sustain a work in existence. This also leads to ontic vagueness, as well as other problems. On the whole, while it is not a perfect view, I accept Thomasson’s position that only a potential reader is required to sustain a work in existence.

Now if proponents of the anti-gappiness strategy could show that there existed potential readers of *The Tale of Sinuhe* and other literary works written in ancient Egyptian hieroglyphics during the period of time from 400 A.D. until 1824 (obviously there were no actual readers of hieroglyphic texts during this time), this would establish that *The Tale of Sinuhe* and other hieroglyphic works existed during that period as well. This would eliminate all supposed gaps in these literary works. *The Tale of Sinuhe* has existed since its creation in 1960 B.C. because there has always existed at least one

---

physical copy of its text and there has always existed at least one person who potentially could read and properly interpret this work.

At first it sounds rather implausible to claim that there was at least one person alive in 500 A.D., 1066, 1492, 1776, and so on who was a potential reader of ancient Egyptian hieroglyphic texts. No person spoke ancient Egyptian by this point in time. No person was writing new texts in ancient Egyptian during this time period. No person actually read any ancient Egyptian hieroglyphics during that time period. A few scholars did attempt to decipher hieroglyphics over the centuries, but none met with success. Surely no educated person from this time period would have considered himself a “potential reader” of ancient Egyptian hieroglyphics.

However the concept of a “potential reader” of a language is not strictly defined and is open to quite a bit of interpretation. Anti-gappiness strategists can offer a fairly plausible pair of necessary and sufficient conditions for a person to qualify as a potential reader of a language. In order for there to be at least one person who is a potential reader of language $l$ at time $t$, there must exist at least one physical copy of some text written in $l$ at $t$, and there must exist at least one physical copy of a text that functions as a crib for $l$ at $t$.

A crib is any text that contains both a known language and an unknown language to be learned, and the crib enables a prospective student of the unknown language to learn it by means of the known language. For instance, in the case of English speakers learning Latin, examples of cribs would include English-Latin dictionaries, Latin grammar books, the Loeb series of side-by-side translations of Latin classics, and the chart my high school Latin teacher wrote on the chalkboard indicating how to conjugate verbs in the perfect
tense. Obviously some cribs are more effective than others in enabling readers to learn the second language.

The Rosetta Stone functioned as a crib text for ancient Egyptian hieroglyphics. The inscriptions on the Rosetta Stone consist of a language that has always been known, ancient Greek, and the unknown language of ancient Egyptian hieroglyphics. Anyone who could read ancient Greek could use the Rosetta Stone to gradually decipher hieroglyphics.

Given that the Rosetta Stone crib existed during the period from 400 A.D. until 1824 and given that many physical copies of hieroglyphic texts existed during that period as well, there also existed at least one potential reader of ancient Egyptian at every moment during this time period. Since there was always at least one potential reader of ancient Egyptian from 400 A.D. to 1824, this means that literary works written in hieroglyphics like *The Tale of Sinuhe* existed without interruption from 400 A.D. to 1824. There are no temporal gaps in *The Tale of Sinuhe*’s “lifespan.”

4.5 Objections to the Anti-Gappiness Strategy

Two objections to the anti-gappiness strategy come to mind. Both objections take issue with the strategy’s claim that the existence of a crib at *t* is both a necessary and sufficient condition for the existence of potential readers of language *l* at *t*. The first objection asserts that the existence of a crib for language *l* is a necessary but not sufficient condition for the existence of potential readers of *l*. The second objection
asserts that the existence of a crib is not a necessary condition for the existence of potential readers of \textit{l}.

The first objection runs along these lines: when considering the relationship between people alive from 400 to 1824 and ancient Egyptian, it is still far-fetched to claim that at least one of those people was a potential reader of ancient Egyptian. The only surviving crib for ancient Egyptian, the Rosetta Stone, was buried and completely inaccessible during that time period. No one alive during those 1400 years even knew that the Rosetta Stone had existed in the first place, let alone that it still survived hidden within an ancient wall. Now it seems that in order for a person to qualify as a potential reader of any text, the text must be reasonably accessible to that person. For example, my husband owns a copy of \textit{Piers Plowman}, a book that I have never actually read. Nonetheless I am a potential reader of this copy of \textit{Piers Plowman} because it sits upon a bookshelf in my home.\footnote{I am also a potential reader of \textit{Piers Plowman} because it is written in English and I know how to read English.} It would be very easy for me to access that literary work. Other works would be more difficult for me to access, but this does not seem to lessen the fact that I am a potential reader of these works. If the sole surviving copy of \textit{Piers Plowman} were kept in a library in Oxford, it seems I would still qualify as a potential reader of this text, even though I would have to travel a great distance to read it. After all, I have the potential to travel to Oxford.

But at a certain point it seems that a text can become so inaccessible to a person that the person no longer qualifies as a potential reader of that text. If the sole surviving copy of \textit{Piers Plowman} were put in a capsule at the bottom of the ocean, or if it were sent
to Mars on a space probe, then it does not seem to be the case that I (or anyone else for that matter) would qualify as potential readers of *Piers Plowman*.

Given the historical facts, it seems that the Rosetta Stone was far too inaccessible from 400 to 1799 for any person alive to qualify as a potential reader of it. The Rosetta Stone’s location within a wall during that time period is more analogous to being on Mars or at the bottom of the sea than it is to being in a library in another country. Since the Rosetta Stone functioned as a crib for ancient Egyptian, if no person from 400 to 1824 was a potential reader of the Rosetta Stone then no person from 400 to 1824 was a potential reader of ancient Egyptian. In that case, literary works like *The Tale of Sinuhe* are once again gappy.

There really is only one way for the anti-gappiness strategist to respond to this objection: simply insist that people alive from 400 to 1799 had the potential to access the Rosetta Stone, and consequently were potential readers of hieroglyphics. The anti-gappiness strategist could argue that the location of the Rosetta Stone from 400 to 1799 is not akin to being on Mars but rather to being in a very far away library. The Rosetta Stone, after all, was buried within an ancient wall and it was discovered when Napoleon’s soldiers demolished the wall in 1799. There was nothing unique about these soldiers and their demolition technology that made it the case that they and only they could have discovered the Rosetta Stone. Countless people who were alive from 400 to 1799 could have demolished that wall and discovered the Rosetta Stone. Given that many people between 400 and 1799 could have torn down the wall and recovered the Rosetta Stone, these people qualified as potential readers of the Rosetta Stone, and thus potential readers of ancient Egyptian.
Whether this is a convincing response on the part of the anti-gappiness proponent is difficult to say. Ultimately these sorts of debates are between two sets of competing intuitions. Some philosophers will have the intuition that the Rosetta Stone, when buried in the wall, was too inaccessible for anyone to qualify as a potential reader of it. Other philosophers will have the opposite intuition. Most people, even philosophers, will probably lack intuitions either way about such a case. It is difficult to see how such debates can be resolved, given that they concern our intuitions about a concept, potential reader, that is already very vague.

The second objection to the anti-gappiness strategy is much more devastating. This objection takes issue with the strategy’s claim that the existence of a crib is a necessary condition for there to be potential readers of a language. At first this claim sounds reasonable; it is hard to imagine how modern scholars could decipher texts in a lost language if there are no cribs for that language. The apparent lack of cribs for Etruscan, to give one example, is why Etruscan is considered a truly dead language and why Etruscan literary works such as the hypothetical poem “Xeqhu” are non-existent despite the existence of physical copies of their texts.

Nonetheless there is a relatively recent instance of a lost language being rediscovered without the aid of any cribs, the decipherment of Linear B. Linear B is an archaic form of pre-Homeric Greek, spoken on Crete roughly between 2800 B.C. and 1100 B.C. Many clay tablets inscribed with Linear B texts have survived from the period, although it is important to note that these tablets almost exclusively describe palace inventories and military maneuvers and are not, strictly speaking, examples of

---

literary works. Now the fact that we know today that these clay tablets are inventories and not poems or plays indicates that Linear B has been translated into living languages and there exist many people who can read texts written in Linear B. But this translation was no small feat, for there is not a single crib for Linear B.\textsuperscript{391} There is no Rosetta Stone-like object that contains a Linear B text alongside a Homeric Greek translation. Complicating matters is the fact that Linear B employed an alphabet composed of fanciful characters that bear absolutely no resemblance to any other alphabet current or ancient. Among other things, characters in the Linear B alphabet resemble:

\[A\] Gothic arch enclosing a vertical line, a ladder, a heart with a stem running though it, a bent trident with a bar, a three-legged dinosaur looking behind him, an A with an extra horizontal bar running through it, a backward S, a tall beer glass, half full, with a bow tied on its rim.\textsuperscript{392}

Tablets containing Linear B texts were first discovered in 1900, and for the next forty years the consensus among archaeologists, classicists, and linguists was that deciphering this language was a hopeless cause. But in the late 1940s some progress started to be made on “cracking the code” of Linear B. A Brooklyn classics professor by the name of Alice Kober began an extensive study of Linear B. Since Kober had no crib for Linear B, only copies of Linear B texts, her approach to deciphering the script had more in common with cryptanalysis than traditional linguistic endeavors like Champollion’s. Kober paid close attention to the pairs and triples of Linear B characters that were most frequently found on the clay tablets. She carefully noted how certain pairs and triples of characters always appeared at the ends of words. From this she deduced that Linear B was an

\textsuperscript{391} Singh, \textit{The Code Book}, 218. \\
\textsuperscript{392} Singh, \textit{The Code Book}, 220-221.
inflected language. Kober also speculated that certain characters in Linear B were vowels while others were consonants, based upon the way these symbols were paired with one another and their positions within words.\textsuperscript{393}

This was groundbreaking research to be sure, but it was not yet a decipherment of Linear B. Kober may have indicated that certain alphabetic characters were vowels or consonants, but she had no idea of which consonants and vowels corresponded to which characters. And unfortunately before Kober could further her research on Linear B she died of cancer in 1950.

Waiting in the wings was a young architect and amateur classicist by the name of Michael Ventris. Ventris had become aware of Kober’s work on Linear B in the late 1940s, and after her death he continued her work of analyzing the frequencies and locations of character pairs and triples.\textsuperscript{394} Ventris was able to deduce whether or not certain Linear B characters were consonants or vowels, and he also deduced that some Linear B characters represented syllables while others represented only single letters. Yet he faced the same problem that Kober faced: knowing that a specific Linear B character is a consonant (or consonant syllable) is not enough, one must know exactly which consonant the character stands for if one is to have any hope of deciphering Linear B.

At this point, Ventris simply followed a hunch he had and speculated that three of the most frequently found words in Linear B were the names of the three largest towns in ancient Crete. Using the known names of these towns, Ventris was able to indicate exactly which sounds specific Linear B characters represented; for example, the

\textsuperscript{394} Singh, \textit{The Code Book}, 231.
beribboned beer glass character represents the syllable “ki.” Eventually this discovery enabled Ventris to determine the specific letter or syllable represented by every character in the Linear B alphabet. Once the letters and syllables used in Linear B were known it also became apparent that Linear B was an archaic form of Homeric Greek. Since Homeric Greek was a known language in the 1950s, linguists were finally able to translate Linear B texts into English.\textsuperscript{395} For instance, to a philologist it is clear that the Linear B word \textit{lewotrokhowoi} is an earlier form of the Homeric Greek word \textit{loutrokhooi}. Since it is already known that \textit{loutrokhooi} means “bath-pourers” in English, it becomes possible to translate the Linear B word \textit{lewotrokhowoi} into English and other living languages.\textsuperscript{396}

4.6 The Anti-Gappiness Response to Linear B

Linear B was completely deciphered without the aid of any cribs. This was a wonderful achievement as far as archaeologists, classicists, and linguists were concerned. But the decipherment of Linear B presents an enormous challenge to the anti-gappiness strategy. The anti-gappiness strategy claims that in order for a literary work $w$ to exist at $t$ there must exist at $t$ at least one person who is a potential reader of $w$. In order to be a potential reader of $w$ one must be able to read the language $l$ that $w$ is written in. So in other words, in order for $w$ to exist at $t$ there must be at least one person who is a potential reader of $l$ at $t$. Now in order for anyone to qualify as a potential reader of $l$ at $t$,

at least two kinds of things must exist: at least one physical copy of a text written in \( l \), and at least one physical copy of a text that functions as a crib for \( l \).

Anti-gappiness proponents must say that there has always been at least one person who was a potential reader of Linear B at every moment between 1100 B.C. and 1955 A.D.\(^{397}\) If anti-gappiness proponents deny this then they will either have to argue that there were no potential readers of Linear B after 1100 B.C. – which is wildly implausible since there have been several actual readers of Linear B since that time period such as Michael Ventris – or they will have to concede that Linear B is temporally gappy - it existed from 2800 B.C. until 1100 B.C. and came back into existence in 1955 A.D. But anti-gappiness proponents cannot convincingly argue that there was always at least one person who qualified as a potential reader of Linear B from 1100 B.C. until 1955 A.D. Such an argument ultimately conflicts with the very heart and soul of artifactualism.

Consider just how an anti-gappiness proponent would argue that people alive between 1100 B.C. and 1955 were potential readers of Linear B. The anti-gappiness proponent would probably try to draw some comparisons between the status of Linear B during this time period and the status of ancient Egyptian from 400 A.D. to 1824. The anti-gappiness proponent believes that some people alive between 400 and 1824 were potential readers of hieroglyphics because physical copies of hieroglyphic texts existed and a crib - the Rosetta Stone - existed. As for Linear B, there were physical copies of Linear B texts between 1100 B.C. and 1955, but there were no cribs during that time period. What seemed to replace a crib when Linear B was actually deciphered was simply the creativity, hard work, and above all the intelligence of Kober and Ventris.

\(^{397}\) I am using 1955 as the date when Linear B came back into existence since it is the year when Ventris and the Cambridge classicist John Chadwick completed their work on *Documents in Mycenaean Greek*, a complete decipherment of Linear B’s alphabet and words (Singh, 242).
The existence of at least one person who is very intelligent and dedicated functions as a crib in the case of Linear B.

Perhaps anti-gappiness proponents should revise their conditions for the continued existence of literary works. A literary work exists at \( t \) as long as at least one person who is a potential reader of the work’s language \( l \) exists at \( t \). A person qualifies as a potential reader of a work’s language at \( t \) as long as there is at least one copy of a text written in \( l \) at \( t \) and that person is smart enough and diligent enough to decipher \( l \) at \( t \). The person need not actually decipher \( l \); he or she simply must have the potential to do so if so inclined.

Thus Linear B has existed without interruption since 1100 B.C. because physical copies of Linear B texts have been in existence since that time and surely there has always been at least one person who was smart enough to decipher Linear B since that time. No intelligent person attempted to actualize this potential to decipher Linear B until the 1940s, but there surely must have existed many intelligent people before the 1940s who had possessed the same level of intelligence, creativity, and dedication that Kober and Ventris did. These people, for whatever reason, did not attempt to decipher Linear B but surely some of them could have, just as no people between 400 and 1799 actually dug up the Rosetta Stone but surely some of them could have.

But now the anti-gappiness strategist has stretched the definition of a “potential reader” of a language to a point where it has become nearly meaningless. A potential reader of any language is simply any intelligent and hard-working person, even if that person does not actually attempt to decipher a lost language or is even aware that said language is waiting to be deciphered. Given this excessively liberal account of a
potential reader, it is difficult to see how any human language (along with the literary works written in it) can ever be destroyed. Etruscan, which I previously put forth as a dead or non-existent language, actually still exists at this moment because there is (presumably) at least one person alive at this moment who is intelligent enough to decipher Etruscan. Perhaps this person is actually trying to decipher Etruscan but has not met with any success yet, or perhaps this person is working in another field entirely. Either way it does not matter; Etruscan exists in 2011 and any literary works written in Etruscan such as “Xeqhu” also exist in 2011.

What is so troubling about this, as far as artifactualists are concerned, is that literary works depend upon the existence of both spatio temporal-entities (copies of their texts) and the existence of specific mental states (being able to read a language and being able to interpret a work properly), in order to continue to exist. If either condition is lacking then the literary work will not exist. Now if artifactualists genuinely believe that literary works depend upon the existence of their copies and the existence of mental states like knowing how to read a language, then they must concede that a literary work can be destroyed via the destruction of all copies of the text, and via the destruction of specific mental states. But the anti-gappiness strategy and its response to cases like Linear B indicates that it is next to impossible to destroy a language. No language can go out of existence as long as at least one intelligent person exists. Thus it is next to impossible for any literary work to go out of existence because its language has become dead. Instead the only state of affairs that can still destroy a literary work is the destruction of every copy of its text.
This violates the very core of artifactualism. If a literary work cannot be destroyed via the destruction of its language (since languages are virtually impossible to destroy), then does it really make sense to say that the literary work depends upon its language being a living one? If the only thing that can destroy a literary work is the destruction of all the copies of its text, then it seems reasonable to conclude that the only thing a literary work depends upon for its continued existence is the continued existence of at least one copy of its text. But this conflicts with the entire artifactualist account of literary works and their ontology. In fact, the claim that a literary work only depends upon the existence of a copy of its text seems to skirt very close to the disreputable physical object hypothesis.

The anti-gappiness strategy simply does not succeed, and given its ultimate implications for artifactualism I doubt that it is worth salvaging. Now this leaves artifactualists back at square one, trying to account for apparently gappy literary works like *The Tale of Sinuhe*. Given all the trouble that comes with trying to eliminate gaps in literary works like this, perhaps it is time to give temporal gappiness a second chance.

4.7 How I Learned to Stop Worrying and Love the Gaps

Let us return our attention to *The Tale of Sinuhe*, a real life example of a literary work that, according to artifactualism, is temporally gappy. *The Tale of Sinuhe* was composed sometime around 1960 B.C., which for the sake of simplicity I will designate as time $t_1$. Around 400 A.D., time $t_4$, the last people who could read texts written in hieroglyphics died, and *The Tale of Sinuhe* went out of existence. In 1824, time $t_{18}$,
Champollion used the Rosetta Stone to decipher hieroglyphics, which brought *The Tale of Sinuhe* and other ancient Egyptian literary works back into existence. Those works continue to exist to this day, time $t_{20}$.

*The Tale of Sinuhe* came into existence at $t_1$, went out of existence at $t_4$, and came back into existence at $t_{18}$. *The Tale of Sinuhe* that exists from $t_1$ to $t_4$ is identical to *The Tale of Sinuhe* that exists from $t_{18}$ to $t_{20}$. Initially I believed that no artifactualist could say that a literary work that exists from $t_1$ until $t_4$ is identical to a literary work that exists from $t_{18}$ until $t_{20}$. According to artifactualism, literary works have their origins essentially. A literary work that originates at $t_1$ can never be identical to a literary work that originates at $t_{18}$. But as I have shown in this chapter, attempts to eliminate this kind of gappiness and show that somehow a work like *The Tale of Sinuhe* exists without interruption from $t_1$ until $t_{20}$ do not succeed.

Up until this point the main reason I have been keen to avoid gappiness in literary works at all costs is my concern that gappiness is incompatible with artifactualism’s insistence upon the essentiality of a work’s origins. But perhaps this concern is misplaced, and a literary work can be gappy while maintaining the essentiality of its origin. It is indeed true that a literary work created at $t_1$ is not identical to a literary work created at $t_{18}$. Even if the two literary works have perfectly identical texts, they are still two distinct literary works because they originated, or were created, at different points in time. But in the case of *The Tale of Sinuhe*, which exists from $t_1$ to $t_4$ and $t_{18}$ to $t_{20}$, is it really the case that a literary work is created at $t_{18}$? The answer to this question is no twice over, so to speak. At $t_{18}$, a literary work – *The Tale of Sinuhe* – comes into
existence, but it is not *created*. At t18 an entity is created, but this entity is not a *literary work*, let alone *The Tale of Sinuhe*.

To understand why *The Tale of Sinuhe* comes into existence at t18 but is not *created* at t18, recall artifactualism’s account of literary creation. A literary work is created when its author deliberately engages in a process known as composition.\(^{398}\) Creating or composing a literary work involves creating a text. Usually the composition of a text results in the creation of a physical object like ink marks on a piece of paper or data encoded into a computer’s hard drive. The composition of a text must be accompanied by specific authorial mental states; the author must *intend* to compose a literary work as he creates his text.\(^{399}\) Artifactualism rules out “swamp Pride and Prejudice” and the like. The result of this creative process is a *literary work* - a short story, novel, poem, or drama - an entity that has “certain aesthetic and artistic qualities and ordinarily telling a tale concerning various characters and events.”\(^{400}\)

This bears little resemblance to what happens to *The Tale of Sinuhe* at t18. At t18 Champollion, armed with his recently acquired knowledge of, reads the text of *The Tale of Sinuhe*. For the sake of simplicity, let us assume that Champollion interprets *The Tale of Sinuhe* correctly. *The Tale of Sinuhe* comes into existence at t18 because at t18 there exists at least one physical copy of its text and at least one person who can read and understand it (Champollion). But this is not the creation of a literary work. In order for *The Tale of Sinuhe* to be created at t18, some act of composition must occur at t18. But no such act occurs; Champollion is reading, but he is not composing any new text.

---

\(^{398}\) Thomasson, *Fiction and Metaphysics*, 64.


\(^{400}\) Thomasson, *Fiction and Metaphysics*, 64.
Furthermore in order for *The Tale of Sinuhe* to be created at $t_{18}$, whoever is created it – Champollion perhaps – must intend to compose a literary work at $t_{18}$. But we have no indication that Champollion was intending to compose anything as he sat and read the papyrus. The notion that he was creating a literary work of his own as he sat and read *The Tale of Sinuhe* would probably have struck Champollion as utterly ridiculous.

Now some one hundred and fifty years later, some of Champollion’s compatriots, such as Roland Barthes, would not have found such a notion ridiculous. Barthes’ distinction between author and scriptor is part of a family of views known as *constructivism*. In the previous chapter, I defined constructivism as the view that individual readings or interpretations of literary works somehow themselves constitute or create numerically distinct literary works. Artifactualism bears some important similarities to constructivism but it is not a constructivist view itself. While constructivism is mistaken (and wildly counter-intuitive at times, to boot) it has one extremely important insight. Some sort of entity is created when a person reads and interprets a literary work. This kind of thing can be called many names: an interpretation, an object of interpretation, a reading, an objection of attention, and so on. This thing, which I will call an interpretation, is a complex mental state or collection of mental states had by the reader that might involve visualizing the events in a literary work and speculating about what ambiguous or complicated parts of the literary work mean.

In the case of *The Tale of Sinuhe*, when Champollion read it for the first time at $t_{18}$, some thing was created at $t_{18}$ – Champollion’s interpretation of *The Tale of Sinuhe*. But clearly Champollion’s interpretation of *The Tale of Sinuhe* – a collection of

---

Champollion’s thoughts, imaginings, and beliefs about the story – is not a literary work. It could not be a literary work, given the conditions required for one to create a literary work. But Champollion’s interpretation of *The Tale of Sinuhe* is a full-fledged object in its own right, and this object came into existence at $t_{18}$. One simply must be careful not to confuse Champollion’s interpretation of *The Tale of Sinuhe* with *The Tale of Sinuhe* itself.

So my earlier concern that gappy literary works would conflict with works’ essentiality of origins is unfounded. There is only a conflict between gappiness and essentiality of origins if a work created at $t_{1}$ is supposed to be identical to a literary work created at $t_{18}$. But closer examination of real life cases of gappy works, such as *The Tale of Sinuhe*, indicates that no literary work is created at $t_{18}$. A literary work comes into existence at $t_{18}$ but it is not created at $t_{18}$ because the conditions for the creation of a literary work are not present at $t_{18}$. Likewise something is created at $t_{18}$, but it that thing is an interpretation, not a literary work itself. Literary works like *The Tale of Sinuhe* can be temporally gappy without compromising the essentiality of their origins.

It should be noted that literary works are not the only entities that are gappy at times. There are many other cultural, social, and political entities that – like literary works – depend upon a combination of spatio-temporal objects and intentional states in order to be instantiated and – like literary works – are temporally gappy at times. Rutgers University was chartered in 1766 but closed in 1795 due to a lack of funds. In 1808 Rutgers was re-opened and exists until the present day. The classic rock band The Who formed in 1964, broke up in 1983, reformed again in 1989, disbanded within a year, and reunited a second time in 1996. The independent nation of Poland existed from 966 A.D.
until 1795 when it was partitioned among three other nations. In 1918 Poland, as an independent country, came into existence again. Ordinary intuitions and practices count one Rutgers, one The Who, and one Poland not two, three, and two (respectively) distinct entities in these cases. While the possibility of temporally gappy persons and spatio-temporal objects strikes us as odd, we seem quite comfortable with gappiness in social entities. In fact the only social entities that come to my mind where ordinary intuitions and practices do not allow for gappiness are marriages. If the same two people marry, divorce, and then remarry at a later date, we seem to want to count this as two distinct marriages, not one marriage with a temporal gap. Note how Elizabeth Taylor is described as having been married “eight times,” with the caveat “twice to the same man” almost always following.

4.8 Concluding Remarks

Given that many familiar entities from everyday life are gappy, acknowledging that literary works can be gappy as well should not be too great of a leap. Despite initial appearances to the contrary, one can reconcile artifactualism’s claim that literary works have their origins essentially with those same works being gappy. Artifactualism’s unique take on the relationship between a work and its interpretations allows artifactualists to embrace gappiness in literary works without compromising the essentiality of those works’ origins.
In this final chapter, I will address some important questions that have been postponed until now. The main question that will be considered here is how incorrect must an interpretation of a literary work be before it can lead to that work’s destruction? This is a version of the much larger question of what constitutes valid or correct interpretation – a very difficult question indeed. I will explore two interesting, although not definitive, answers to the question of valid interpretation and discuss their various merits and flaws.

I also will consider the practical consequences of artifactualism and its claims about interpretation. The destructive potential of misinterpretation is not a merely hypothetical scenario used to bolster some arcane metaphysical claim. Rather it is something that can and does happen, with literary works originating from oral tradition being particularly vulnerable to destruction via misinterpretation.

5.1 Interpretation and Misinterpretation

According to Amie Thomasson’s account of artifactualism, a literary work $w$ exists at time $t$ if and only if at least one copy of $w$’s text exists at $t$ and at least one person with “the language capacities and background assumptions” required to “read and
In Chapter 3 I argued that the requirement that a reader have the proper “background assumptions” is best understood as requiring the reader to be able to properly interpret the work. Therefore, according to this revised version of artifactualism, a literary work \( w \) exists at time \( t \) if and only if at least one copy of \( w \)’s text exists at \( t \), and at least one person who can read the language \( w \) is written in and who can properly interpret \( w \) exists at \( t \). Literary works depend upon being properly interpreted in order to enjoy continued existence. If for some reason there is not a single person alive on earth who can properly interpret a literary work, then that work no longer exists – even if physical copies of its text remain in existence.

To interpret a work is to say or think something about the meaning and significance of that work. As I explained in Chapter 3, I have purposely left this account of interpretation on the under-defined side, given that “interpretation” commonly denotes an extremely broad and diverse range of practices. What all of these different practices, engaged in by different types of readers for different purposes, seem to have in common is saying something about the meaning and significance of literary works.

Interpretation can refer to both the practice of interpretation and to the object that is produced by that activity. When I discuss interpretation qua object – which I often do in this project – such an object is a mental state or collection of mental states produced by

---


a reader that aims to describe the meaning and significance of a literary work (or of a part of a literary work).

Artifactualism is not simply concerned with interpretations, but specifically with correct and incorrect, or proper and improper, interpretations. A literary work depends upon there being at least one reader who can not only interpret it but who can properly interpret it. I shall define a proper interpretation as an interpretation that provides a correct account of the meaning and significance of a literary work. An improper interpretation – or a misinterpretation – is an interpretation that provides a false account of the meaning and significance of a literary work.

It must be noted that by giving these definitions of interpretation and misinterpretation, I am committing myself to the view that there are standards of correctness for interpretation and that interpretation is capable of reaching objective truth. This is not an uncontroversial position. Since the mid-twentieth century, there have been several highly influential theories of criticism ranging from Hans-Georg Gadamer’s radical historicism,⁴⁰⁴ to the deconstructionism of Jacques Derrida and Michel Foucault,⁴⁰⁵ to the reader response criticism of Stanley Fish,⁴⁰⁶ that all express skepticism if not outright denial that literary interpretation can arrive at objective truth. According to these theorists, there is no meaning and significance of a literary work independent of a reader’s personal interpretation of it. A work simply means “what it means to us.

---

⁴⁰⁶ Stanley Fish, “Interpreting the Variorum,” 152, 159, 159.
today.\textsuperscript{407} Under this skeptical conception of interpretation, it is incoherent to speak of correct and incorrect interpretations.

It is beyond the scope of my project to offer a sustained refutation of interpretative skepticism. I do wish to note that such skepticism has many undesirable consequences for both literary studies as well as the other humanities. Denial of literary interpretation’s ability to arrive at objective truth ultimately leads to a denial of any humane science to arrive at objective truth – to arrive at knowledge. E.D. Hirsch notes:

Since all humane studies . . . are founded upon the interpretation of texts, valid interpretation is crucial to the validity of all subsequent inferences in those studies. The theoretical aim of a genuine discipline . . . is the attainment of truth, and its practical aim is agreement that truth has probably been achieved. Thus the practical goal of every genuine discipline is consensus – the winning of firmly grounded agreement that one set of conclusions is more probable than others – and this is precisely the goal of valid interpretation.\textsuperscript{408}

Going forward, I will take for granted the position that the practice of interpretation can, at least in principle, enable readers to reach objective truth about literature. For any literary work, there are facts about what its meaning and significance are. Interpretation is the practice by which a reader discerns those facts. Because the goal of interpretation is the discovery of facts about what a work of literature is like, both proper interpretation and misinterpretation are possible.

If one rejects interpretative skepticism, it will be quite obvious that proper interpretations are to be sought after and misinterpretations avoided. Proper interpretations enable readers to discover important facts about literary works, thus attaining knowledge about literary works. Meanwhile misinterpretations impede the

\textsuperscript{407} Hirsch, \textit{Validity in Interpretation}, viii.
\textsuperscript{408} Hirsch, \textit{Validity in Interpretation}, viii-ix
discovery of important facts about literary works, hindering the achievement of knowledge. But for the artifactualist, misinterpretation not only presents an epistemic problem but also an ontological problem. Under certain circumstances, the misinterpretation of a literary work will result in the destruction of that work.

The certain circumstances I have in mind here should be familiar from Chapter 3. A work of literature can be destroyed via misinterpretation when something like the following happens: the relevant background information needed to properly interpret the work is lost to the entire community of potential readers. Not one single person alive has the ability to properly interpret the work. But nonetheless people still read the work and still attempt to interpret it. Because, unbeknownst to them, they lack the relevant information needed to properly interpret the work, all their interpretative activity produces misinterpretations of the work. These misinterpretations are not capable of sustaining the work in existence; the work is destroyed via misinterpretation.

I will call these sorts of misinterpretations “potentially destructive.” They are potentially destructive – not destructive simpliciter – because it is only under very specific circumstances, where no person can supply the proper interpretation of a given literary work, that such misinterpretations become destructive. Most of the time when misinterpretation occurs it happens as an isolated incident; a single reader or small group of readers makes an egregious mistake concerning the meaning and significance of a literary work, such as believing *Animal Farm* is a strange children’s story. These familiar acts of misinterpretation are not destructive because, in nearly every case, the mistaken readers will be greatly outnumbered by people who can properly interpret the work in question. The presence of many people who can properly interpret *Animal Farm,* for
instance, is what enables *Animal Farm* to continue to exist, despite a few misinterpretations of it here and there.

A very important question naturally arises from these claims: at what point does a misinterpretation of a literary work become potentially destructive? In other words, how wrong must a misinterpretation be before it can, under certain circumstances, destroy the work? It will not do to claim that *every* misinterpretation of a work is potentially destructive of that work. Some misinterpretations are quite minor. For example in Yeats’ “The Second Coming” there is a reference to “mere anarchy” being “loosed upon the world.” Almost every person who reads this poem today reads “mere anarchy” as meaning something along the lines of “just anarchy” or “simply anarchy.” This is understandable given the modern connotations of the word “mere.” But in 1920s Irish usage, the connotations of “mere” were different, and “mere anarchy” meant something more like “total anarchy” or “nothing but anarchy.” Few modern readers are aware of this and, as a result, most people who read “The Second Coming” misinterpret it, because most people make an error regarding the meaning and significance of part of the poem.

But surely this is a very minor and inconsequential misinterpretation of “The Second Coming.” Given that the correct information about the original connotations of “mere” is rather obscure, it is not difficult to imagine a world where everyone has forgotten this bit of information and everyone reads “mere anarchy” as meaning “just anarchy.” Everyone misinterprets “The Second Coming” in such a world. Yet it strikes me as radically counter-intuitive to claim that “The Second Coming” would no longer exist in this world. Perhaps if everyone failed to recognize that “The Second Coming” had an apocalyptic theme, or failed to recognize that it even was a poem to begin with,
then “The Second Coming” would no longer exist. But a mistaken belief about the
connotation of one word in a poem does not seem to be a potentially destructive
misinterpretation.

Many instances of misinterpretation seem similarly minor and ontologically
benign. For any given work of literature, there are countless facts about it, properties,
and features of it and its parts that fall under the umbrella of the work’s “meaning and
significance.” Surely many, if not most, readers will fail to recognize the presence of all
these facts, properties, and features. A reader might fail to note an allusion to another
work, misunderstand a metaphor, not know that a character was inspired by the author’s
sister, be ignorant of how two characters are foils for one another, overlook a minor plot
point, misjudge the connotation of a word, and so on. These all seem to me to be
examples of minor misinterpretations. Major misinterpretations – potentially destructive
misinterpretations – seem to me to involve serious interpretative mistake such as failing
to recognize a work’s genre, failing to recognize a work’s theme, misunderstanding major
plot points, grossly misjudging the nature of the main characters, failing to pick up on
centrally important allusions, metaphors, or symbolism, and so on.

Yet simply stating what sorts of misinterpretations seem to me to be minor or
potentially destructive is not particularly useful. While it is helpful to consult one’s
intuitions about this matter, doing so does little to answer the question of what
distinguishes potentially destructive misinterpretations from minor misinterpretations. In
fact, a wholesale examination of one’s intuitions about misinterpretation raises more
questions than it answers. There is a great deal of vagueness about which sorts of
properties and features of literary works are centrally important for correct interpretation
and which are not. For instance, if misunderstanding a work’s major plotlines is potentially destructive while misunderstanding a work’s minor plotlines is benign, how are we to distinguish between the major and minor plotlines of a given work? If a metaphor is found in a work, how are we to determine whether it is centrally important to the work or not?

In the following sections, I will explore ways of providing a more definitive answer to the question of how wrong a misinterpretation must be before it becomes potentially destructive. The question of how to distinguish potentially destructive misinterpretations from minor misinterpretations is actually a twist to the classic question of how to establish standards of correctness for interpretation. The question of what constitutes correct interpretation is an enormously important question – sometimes called the central question of the humanities – and it is enormously difficult to answer. It is well beyond the scope of this project to answer this question, and I in no way presume to provide a definitive answer to it here. My aim in the following three sections is quite modest; I will explore two approaches to the question of potentially destructive misinterpretation, approaches that I consider relatively promising.

5.2 Literary Works and Essential Properties

My first suggestion for establishing how incorrect a misinterpretation may be before it becomes potentially destructive is to appeal to a distinction between the essential and accidental properties of literary works. I will call this view essentialism about literary works, or simply essentialism. The essentialist picture of literary works is
roughly this: literary works are objects, and just like spatio-temporal objects, literary works have properties. For any given literary work and its properties, some of those properties will be accidental to the work and some of those properties will be essential to it. Literary works have individual essences – there are essential properties that *Pride and Prejudice* and only *Pride and Prejudice* has, there are essential properties that *Ulysses* and only *Ulysses* has, and so on. Finally the set of a literary work’s essential properties is not limited to trivial essential properties such as being self-identical, but also will include many non-trivial properties – for instance, being a satire is an essential property of *Animal Farm*.

Applying essentialism to the question of misinterpretation can yield a set of principles for distinguishing potentially destructive misinterpretations from correct interpretations and from minor, ontologically benign misinterpretations. An interpretation of literary work $w$ is correct to the extent that it recognizes and acknowledges the presence of $w$’s essential and accidental properties. An interpretation of $w$ is incorrect, but incorrect in a way that is minor and benign, to the extent that it recognizes the presence of $w$’s essential properties but fails to recognize at least one of $w$’s accidental properties. An interpretation of $w$ is a potentially destructive misinterpretation to the extent that it fails to recognize the essential properties of $w$. Of course these principles raise a further question: for a given work of literature, which of its properties are essential and which of its properties are accidental?

---

409 This is not meant to exclude the possibility of shared essential properties. Two literary works can both have the same essential property: for example, both *Pride and Prejudice* and *Emma* have the essential property of being created by Jane Austen.
Thomasson argues that every work of literature has its author and time of creation essentially.\textsuperscript{410} *Portrait of the Artist as a Young Man* essentially has the property of being authored by James Joyce in 1916. Had another author by some Borges-esque coincidence produced a text that was word-for-word identical to *Portrait of the Artist as a Young Man*, the resulting work would be numerically distinct from Joyce’s novel.\textsuperscript{411} Furthermore if Joyce himself had composed a work with exactly the same text as *Portrait of the Artist* but done so in 1922 rather than in 1916, that work also would have been numerically distinct from the novel Joyce wrote in 1916.\textsuperscript{412}

I see no reason to reject Thomasson’s view that a literary work’s author and time of origin are essential to it. But positing these properties as essential to literary works is not very useful for answering questions of correct and incorrect interpretation. Under the essentialist principles outlined above, a correct interpretation of work \( w \) must recognize \( w \)’s author and time of origin. An interpretation of \( w \) that fails to recognize \( w \)’s author and time of origin will be potentially destructive. But is this plausible in light of our ordinary intuitions and practices surrounding interpretation?

It certainly seems to be the case some of the time that in order to properly interpret a work, a reader must know when the work was composed. That *The Man of Mode* and *The Country Wife* were written during the Restoration, or that *A Farewell to Arms* was written in the aftermath of World War I seem to be essential pieces of information to have if one is to successfully interpret those works. Somewhat less certain is the claim that in order to properly interpret a work, a reader must know its author. Certainly it is important in some cases to know a work’s author; an interpretation of

\textsuperscript{410} Thomasson, *Fiction and Metaphysics*, 8-9, 36, 132.
\textsuperscript{411} Thomasson, *Fiction and Metaphysics*, 8, 56-57.
\textsuperscript{412} Thomasson, *Fiction and Metaphysics*, 8.
Sonnet 18 by a reader who knew nothing of Shakespeare strikes me, at least, as seriously impoverished.

But there are far more cases where knowing the author and time of origin of a work does not seem to be necessary for proper interpretation of the work. Presumably, we properly interpret many folktales even though the identity of their authors and their times of origin are shrouded in mystery. It is also hard to imagine why a reader must know the author and time of origin of literary works like romance novels, detective stories, pot-boiler thrillers, and other forms of popular fiction in order to interpret those works properly.

Meanwhile, it is also obvious that knowing the author and time of origin of a literary work is not a sufficient condition for properly interpreting the work. A reader can know, for instance, that *Animal Farm* was written by an author named George Orwell in 1945 – the information is available on the title page, after all – yet still radically misinterpret *Animal Farm* as a strange children’s story.

If the distinction between correct and potentially destructive interpretations is to rest upon the acknowledgement of a work’s essential properties, said essential properties will have to include other properties besides a work’s author and time of origin. Other properties of a work, which one may wish to call the work’s *aesthetic properties*, must make up the work’s essence. Thomasson gives several examples of aesthetic properties that seem to be essential to literary works, although she calls these properties “central” not “essential.” *Animal Farm* is a satire of Stalinism.\(^{413}\) *Portrait of the Artist as a Young Man* uses language in an original, ground-breaking way.\(^{414}\) The screenplay of Oliver

\(^{413}\) Thomasson, *Fiction and Metaphysics*, 8-9.
\(^{414}\) Thomasson, *Fiction and Metaphysics*, 8.
Stone’s *Nixon* is about Nixon, revisionary, speculative, and sympathetic to the protagonist. Mersault’s killing of the Arab is a central feature of *The Stranger*. Other aesthetic properties she considers “central” to works include “being a work of high modernism, a parody, horrifying, reactionary, exquisitely detailed,” and “an updated retelling of an old story.”

Thomasson does not consider the non-essential, accidental properties of literary works much. She does suggest some examples of aesthetic properties that seem to be accidental, in that the literary works that have these properties could survive the loss of these properties. For instance, in a Sherlock Holmes story, Watson is said to have six cuts on his shoe. Had the story been written in such a way that Watson had five cuts on his shoe, it still would have been the same story.

Examples of specific works’ essential and non-essential properties are interesting, but this piecemeal strategy is dissatisfying. Ideally we could provide some universal principles that apply to all literary works for determining which of a work’s properties are essential and which are not. What I have in mind here are principles along the lines of “whenever a literary work *w* has property *p*, *p* is an essential property of *w*” and the like. Unfortunately as one surveys our intuitions and practices surrounding literature and interpretation, the chances of discovering such universal principles appear quite dim. Saying that there are properties that are always essential to literary works – major plotlines, theme, and the like – generates further difficult questions such as how to distinguish a major plotline from a minor one. Meanwhile other properties are essential.

---

to some works and non-essential to others. Consider the property of allusion. The
allusions to the *Odyssey* in *Ulysses* strike me as essential properties of *Ulysses*; it is very
difficult to imagine how a reader could properly interpret *Ulysses* without acknowledging
these allusions. On the other hand, the allusion to *Romeo and Juliet* in Joyce’s *The Dead*
(a painting of the garden scene hangs in the Morkan sisters’ house), does not appear to be
an essential property of *The Dead*. A reader could fail to note the significance of the
allusion (there actually is a very interesting interpretation attached to it) but still properly
interpret *The Dead*.

Instead of discovering universal principles for distinguishing between a literary
work’s essential and accidental properties, we seem to be making such decisions on an
individual, case-by-case basis. The only factor that appears to determine whether or not a
property is judged to be essential or accidental to a given literary work is whether that
judgment is in line with the general consensus among ordinary readers, professional
scholars, and critics. This is not as hopeless an impasse as it may seem, however. The
question of literary works, their essences, and how those essences are determined has
many parallels to the question of the definition of art – a question that led to, among other
theories, the development of the institutional theory of art.

5.3 An Institutional Theory of Literature?

One of the most central and enduring questions within aesthetics is that of the
definition of art. What features distinguish works of art from other non-artistic objects?
Are there necessary and sufficient conditions for an object’s being art? By the mid-
twentieth century the definition of art had become a very pertinent question, given the proliferation of “works of art” that appeared to be completely indistinguishable from ordinary non-artistic objects. Why, for example, was Andy Warhol’s *Brillo Box* considered a work of art while a qualitatively identical object in a grocery store stockroom was not considered a work of art? Older theories of art, such as the mimetic theory, that attempted to provide necessary and sufficient conditions for what made an object art had become inadequate in light of works by Warhol, Rauschenberg, Duchamp, and others. On the practical level, decisions about which objects were art (and which were not) were being made on an individual case-by-case basis without much appeal to universal principles such as art is mimetic, art is expressive, and the like. Rather the determining factor that led to an object’s being considered art was whether or not there was broad consensus among various persons and institutions that the object was art.

The institutional theory of art was developed in response to this situation, most notably by George Dickie. According to the institutional theory of art, an object is a work of art if the art world considers it to be a work of art. The art world is an informal, holistic, and self-defining network of people and institutions such as artists, scholars, audience members, curators, museums, galleries, conservatories, theater

---

419 Some of the more famous examples of this phenomenon include Marcel Duchamp’s sculpture *Fountain*, which is a salvaged urinal displayed on a pedestal, Andy Warhol’s *Brillo Box* which is a balsa wood box hand-painted by Warhol to an exact replica of a commercial Brillo pad container, and the bed sculptures of Claes Oldenburg and Robert Rauschenberg, which are just that – beds. Arthur Danto provides excellent discussion of these and similar artworks and their impact upon the definition of art. See Danto, “The Artworld,” in *Aesthetics: A Critical Anthology*, 2nd ed., eds. George Dickie, Richard Sclafani, and Ronald Roblin (New York: St. Martin’s Press, 1989), 174-180; Danto, *The Transfiguration of the Commonplace*, 5, 12-13, 33-53, 90-91, 94-95, 100, 107-111, 120-123, 132-133, 136-139, 142, 144, 208.


companies, and the like. Armed with these criteria, one can explain why Warhol’s *Brillo Box* is art and the identical object in the stockroom is not art. The former object is considered to be art by the art world; the latter object is not.

This approach to the broad question of the definition of art can be applied to the narrower question of literary works and their essences. Just as there are no universal principles that determine whether a given Brillo box is a work of art, there are no universal principles that determine whether a given property, such as allusion, will be essential or accidental to a particular literary work. Instead, just as judgments about the artistic status of Brillo boxes are made on an individual box-by-box basis, judgments about the essentiality of allusion are made on an individual work-by-work basis. And just as the art world ultimately decides which Brillo boxes are art and which are not, the literary world decides which instances of allusion are essential to their works and which instances of allusion are accidental to their works. Paralleling the account of the art world, the literary world can also be defined as an informal, holistic, self-defining network of people and institutions – authors, readers, critics, scholars, libraries, literary magazines, English departments, and the like.

Ideally, an “institutional theory” of literature would enable us to explain why the allusions to the *Odyssey* in *Ulysses* are essential properties of *Ulysses* while the allusion to *Romeo and Juliet* in *The Dead* is an accidental property of *The Dead*. The literary world considers the allusions in *Ulysses* to be essential, and the literary world considers the allusion in *The Dead* to be accidental. Whenever a reader is puzzled as to whether or not a given property of a literary work is essential to it, he can look to the literary world for the answer. This would also enable one to determine whether or not a

---

misinterpretation of a work was potentially destructive. A misinterpretation of work $w$ is potentially destructive if it fails to recognize $w$’s essential properties. The essential properties of $w$ are those properties that the literary world considers to be essential to $w$.

But note that this is all prefaced by “ideally.” While I am strongly sympathetic to the institutional theory of art in general, a similar approach to the essential and accidental properties of literary works is not going to be as successful. Relying on the literary world to determine which properties of a work are essential would be a sound approach if one typically found broad consensus within the literary world about such matters. Unfortunately even a passing familiarity with the real-life workings of professional scholars and critics of literature indicates that such consensus is quite rare indeed.

One of the first problems with appealing to the literary world to discover which of a work’s properties are essential to it is the fact that, for the most part, few professional scholars and critics of literature use the language of essential and accidental properties when describing literary works. Such terminology is quite foreign to most of the people who make up the core of the literary world. For instance, consider the critic Harold Bloom’s remarks on *The Ballad of Reading Gaol*.

Oscar Wilde, who knew he had failed as a poet because he lacked strength to overcome his anxiety of influence, knew also the darker truths concerning influence. *The Ballad of Reading Gaol* becomes an embarrassment to read, directly one recognizes that every luster it exhibits is reflected from *The Rime of the Ancient Mariner*; and Wilde’s lyrics anthologize the whole of English High Romanticism.\(^\text{425}\)

Bloom notes several of *The Ballad of Reading Gaol*’s properties. It has the property of being influenced by *The Rime of the Ancient Mariner*, the property of being influenced by English High Romanticism, the property of being lustrous, and the property of being a failure in light of Wilde’s views on influence. Yet nowhere in Bloom’s discussion of Wilde is there any indication that Bloom sees, and expects others to see, some of these properties as essential properties of *The Ballad of Reading Gaol*. This is almost always the case in literary criticism. And this presents a serious problem if we are to appeal to the literary world to decide whether or not a property of a work is essential or not. Almost no one in the literary world distinguishes between the properties of literary works in this manner.

To make matters worse, when the literary world has engaged with issues surrounding essences and accidents, it has almost always been in the context of rejecting the entire notion of essentialism. The wholesale rejection of essences – for literary works, for human beings, for anything – is a central tenet of deconstructionism. Anti-essentialist views have also become foundational for post-structuralism, particularly in gender studies and queer theory. Now it is certainly true that not every person in the literary world is an adherent of deconstructionism or post-structuralism; in fact, these theories have sharply declined in popularity in recent years. Nonetheless the influence of these theories, both in the past and to this day, has been very significant as far as the literary world is concerned. The most relevant point for my project is this: any plausible notion of the literary world will have to include as part of that world figures such as

---

426 Richter, “Structuralism, Semiotics, and Deconstruction,” 827.
Foucault, Derrida, and Judith Butler. These critics and their adherents are deeply committed to the rejection of essential properties of any sort. This presents obvious and serious difficulties for those who wish to look towards the literary world for answers to questions like, “is being influenced by The Rime of the Ancient Mariner an essential property of The Ballad of Reading Gaol?”

Fortunately for the essentialist project, there are other schools of thought within the literary world that could be viewed as making claims about which properties of works are essential and which are not. Specifically, I have in mind debates over what sorts of properties and features of literary works interpretation ought to be concerned with. Although it is not phrased this way by critics and scholars, these sorts of disputes can be seen as attempts to determine which properties are essential to proper interpretations of works and which are not.

Consider the evolution of criticism since the mid-nineteenth century. During the nineteenth century a view known as expressivist criticism was dominant. Expressivist critics saw literature as fundamentally a mode of self-expression. A novel, poem, or play was – at its most basic level – its author’s revelations about his or her thoughts, emotions, opinions, and so on. This view was roundly rejected in the mid-twentieth century; the most famous refutation of it is found in Wimsatt and Beardsley’s “The Intentional Fallacy.”

Formalist criticism became the dominant theory of interpretation. Formalist interpretation is characterized by a focus on “the work itself,” careful analysis of the internal structure of the work with very little concern for the author’s biography,

---

Formalism is still very popular among professional critics, but is now rivaled in influence by what is called New Historicism. New Historicism is in many ways an updated version of expressivism. It claims, once again, that an interpreter must be aware of important events in the author’s biography, as well as the historical and cultural circumstances surrounding the creation of a work.

Now imagine, for example, a reader who is attempting to interpret *The Glass Menagerie*. The reader is aware that this play was influenced by Tennessee Williams’ real life relationship with his sister. Being influenced by Williams’ relationship with his sister is a property of *The Glass Menagerie*; there appears to be little controversy within the literary world over this fact. Our reader is keen to properly interpret *The Glass Menagerie* and wonders whether the property of being influenced by Williams’ sister is an essential or accidental property of the play.

Will an appeal to the literary world answer this question? Unfortunately the reader will find sharply divergent answers within the literary world. Formalist critics, while not denying that *The Glass Menagerie* has the property of being influenced by Williams’ sister, will almost certainly deny that this property is essential to a proper interpretation of the work. Meanwhile New Historicists will take the opposite view, claiming that such a property most certainly is essential to any proper interpretation of *The Glass Menagerie*.

At least in the case of *The Glass Menagerie*, there is one area of agreement between the competing Historicist and Formalist interpretations: *The Glass Menagerie*

---

has the property of being influenced by Williams’ sister. But such consensus, as modest as it is, is not the norm in most critical debates. One of the greatest problems with relying on the literary world to determine whether or not a property of a literary work is essential or accidental is that there is often little consensus among those in the literary world as to whether the work even has the property to begin with. A well-known case of this sort of critical disagreement concerns Wordsworth’s poem, “A slumber did my spirit seal.” The last stanza, in which the speaker reflects on the death of his beloved Lucy, has attracted the most critical attention and dispute.

No motion has she now, no force;
She neither hears nor sees,
Rolled round in earth’s diurnal course
With rocks, and stones, and trees.⁴³²

Noted critic Cleanth Brooks interprets this last stanza as meaning Lucy’s death is a dreadful state of affairs; she is “chained” to the earth and caught up in the “violent” whirling motion of the planet. Brooks interprets “A slumber” as taking a bitter and pessimistic view towards death.⁴³³ But another notable critic, F.W. Bateson, interprets the poem quite differently. He reads the same stanza as meaning Lucy’s death is imbued with “pantheistic magnificence.” Bateson interprets “A slumber” as taking a hopeful and awed view towards death. “Lucy is actually more alive now that she is dead, because she is now part of the life of Nature and not just a human ‘thing.’”⁴³⁴

---

Now imagine that a reader is trying to interpret “A slumber” and wishes to avoid potentially destructive misinterpretations. The reader wonders whether the poem’s property of taking a pessimistic attitude towards death is an essential property. Looking to the literary world for guidance on this matter will be fruitless, given that a significant part of the literary world that is familiar with “A slumber” – F.W. Bateson and company – denies that “A slumber” even has the property of taking a pessimistic attitude towards death. The case of Brooks, Bateson, and “A slumber” is well known, but there are countless other examples like this. The problem they pose for essentialism is quite simple – one cannot look to the literary world for consensus about whether a given property $p$ of work $w$ is essential to $w$ if there is no consensus in the literary world as to whether or not $w$ even has $p$ in the first place.

In light of these many problems, appealing to a literary work’s essential properties as a way to define potentially destructive misinterpretations does not appear to be successful. Despite this setback, I remain intrigued by the possibility that literary works have individual essences. I believe it is very much worth exploring the question of whether or not literary works, as well as other types of contingent abstracta – particularly cultural entities like universities, nations, companies, and institutions – have individual essences. Philosophical treatment of this matter could shed more light on our ordinary intuitions and practices surrounding such entities, especially our beliefs about how these entities can and cannot survive changes over time. But defending essentialism about literature and other cultural entities to this extent is clearly beyond the scope of this
project. For now I will set this matter aside and consider another way to characterize potentially destructive misinterpretations.

5.4 Literary Works and Genre

The second way to distinguish between correct interpretations, benign misinterpretations, and potentially destructive misinterpretations of a literary work is by way of an appeal to the work’s genre. Most literary works fall under some sort of genre category. There are the very broad genre categories of novel, short story, play, and poetry; more numerous are narrower genre categories such as tragedy, comedy, drama, epic poetry, lyric poetry, satire, fable, parable, mystery, horror, romance, absurdist, Western, historical fiction, science fiction, and so on. Using genre as a guiding principle, we can formulate standards for distinguishing between correct interpretations, benign misinterpretations, and potentially destructive misinterpretations. An interpretation of literary work \( w \) is correct or benignly incorrect if and only if the interpretation correctly identifies \( w \)’s genre. An interpretation of \( w \) is potentially destructive if and only if the interpretation fails to correctly identify \( w \)’s genre.

Why might anyone think this? Speaking for myself, I came to consider genre especially relevant to correct interpretation as I searched for examples of non-controversially “flagrant” misinterpretations of literary works. When trying to provide examples of indisputably false interpretations that could lead to the destruction of literary works, I found myself time and time again focusing on cases like *Animal Farm* and the dystopian scenario. The misinterpretation that destroys *Animal Farm* in the dystopian
scenario is a genre mistake. *Animal Farm* belongs to the genre of satire, but every reader who attempts to interpret *Animal Farm* in the dystopian scenario fails to identify this genre. This mistake is so serious that it leads to the destruction of *Animal Farm* itself.

For what it is worth, I have the strong intuition that correctly identifying the genre of a literary work is immensely important if one hopes to properly interpret that work. Underlying this intuition is my understanding that interpretation is ultimately about recognizing properties of literary works – properties that fall under the umbrella terms “meaning and significance.” A proper interpretation of literary work $w$ recognizes the properties that $w$ has, while a misinterpretation of $w$ either fails to recognize the properties $w$ has or attributes properties to $w$ that $w$ does not, in fact, have. But in order for a would-be interpreter to recognize the various properties of $w$, it appears that there is one property of $w$ that she must first know – $w$’s genre. One’s beliefs about a work’s genre guide one’s subsequent beliefs about the work and its characteristics in a unique way. When a reader knows that a work is a member of a certain genre, she will bring to her reading of the work a host of expectations and beliefs. She will expect certain properties to appear in the work and other properties not to appear in the work. She may notice properties of the work that she would have otherwise overlooked; for example, if a reader knows *The Glass Menagerie* belongs to the genre of Southern Gothic, she might be more attentive to instances of the grotesque in the play than she would if she knew nothing of the work’s genre beyond “drama.” A reader’s beliefs about a work’s genre also govern her subsequent beliefs about the significance of parts of a work in relation to the whole. An instance of the grotesque means something quite different in a Southern
Gothic novel from what it means in a horror novel and from what it means in a comedic novel.

This “governing power” of one’s genre conception has been noted by other philosophers as well. In “Categories of Art,” Kendall Walton discusses how our perception of a painting or piece of music as belonging to a certain genre category enables us to subsequently perceive many other important properties of that work.\(^ {435}\)

On the other hand, when one mistakenly assigns a work of art to a category it does not belong in, one will quickly develop false beliefs about the properties of that work. Walton asks us to imagine a person exposed for the first time to a cubist painting or a piece of serial music. If this person knows nothing of the genres of cubism or post-tonal music, and instead assumes these objects belong to more familiar genres such as realist painting and Romantic music, she will probably consider these entities to be “formless, incoherent, or disturbing.”\(^ {436}\) However if this person becomes familiar with these new categories, or genres, of painting and music she will soon retract her judgments. She will realize that a cubist painting is not actually formless; there is a specific sort of form unique to cubist works. She will realize the serial music is not actually incoherent; there is a specific sort of coherence unique to serial music. Her mistaken beliefs about the properties of the painting and music were due to a mistaken belief about their genres, and her correct beliefs about the properties of the painting and music are due to a correct belief about their genres. Walton writes, “at least in some cases, it is correct to perceive a work in certain categories, and incorrect to perceive it in certain others; that is, our


\(^ {436}\) Walton, “Categories of Art,” 356.
judgments of it when we perceive it in the former are likely to be true, and those we make when perceiving it in the latter false."\(^{437}\)

One of the most striking examples of the governing power of genre for literary interpretation is a real-life incident recounted by Stanley Fish.\(^{438}\) During the summer of 1971, Fish was teaching two college courses: one focused on linguistics and literary criticism, the other on seventeenth century religious poetry. Both courses met in the same classroom with the poetry course immediately following the linguistics course.\(^{439}\) One morning, Fish wrote the following on the chalkboard:

\begin{center}
Jacobs-Rosenbaum
Levin
Thorne
Hayes
Ohman (?)\(^{440}\)
\end{center}

This was a list of the reading assignments for Fish’s linguistics class. “Jacobs-Rosenbaum” referred to a paper co-authored by Roderick Jacobs and Peter Rosenbaum. The question mark after “Ohman” indicated Fish’s uncertainty about how to spell the last name of Richard Ohmann. Fish left the reading list up on the chalkboard for his poetry course to see. He told the students that this list was actually a religious poem and asked them to interpret it. Armed with a mistaken belief about this text’s genre, the students reacted predictably, providing dazzlingly complex and fascinating, yet utterly false, interpretations of the “poem.”

\(^{437}\) Walton, “Categories of Art,” 356.
\(^{438}\) Stanley Fish, “How To Recognize a Poem When You See One,” in Is There a Text in this Class? The Authority of Interpretive Communities (Cambridge: Harvard University Press, 1980).
\(^{439}\) Fish, “How To Recognize a Poem,” 322.
\(^{440}\) Fish, “How To Recognize a Poem,” 323.
Jacobs was explicated as a reference to Jacob’s ladder, traditionally allegorized as a figure for the Christian ascent to heaven. In this poem, however, the means of ascent is not a ladder but a tree, a rose tree or rosenbaum. This was seen to be an obvious reference to the Virgin Mary who was often characterized as arose without thorns, itself an emblem of the immaculate conception. At this point the poem appeared to the students to be operating in the familiar manner of an iconographic riddle. It at once posed the question, “How is it that a man can climb to heaven by means of a rose tree?” and directed the reader to the inevitable answer: by the fruit of that tree, the fruit of Mary’s womb, Jesus. Once this interpretation was established it received support from, and conferred significance on, the word “thorne,” which could only be an allusion to the crown of thorns. It was only a short step (really no step at all) from this insight to the recognition of Levin as a double reference, first to the tribe of Levi and second to the unleavened bread carried by the children of Israel on their exodus from Egypt. The final word of the poem was given at least three complementary readings: it could be “omen,” especially since so much of the poem is concerned with foreshadowing and prophecy; it could be Oh Man, since it is man’s story as it intersects with the divine plan that is the poem’s subject; and it could, of course, be simply “amen”.

Because so many of a reader’s beliefs about the meaning and significance of a work flow from her beliefs about the genre of that work, a mistaken belief about the work’s genre – as we see in the case of Fish’s “poem” – can snowball into a plethora of mistaken beliefs about the work’s meaning and significance. The sorts of false beliefs about literary works caused by genre misidentifications are typically both serious and numerous. This seems to me to be more characteristic of potentially destructive misinterpretations than benign misinterpretations; therefore a plausible way to distinguish between potentially destructive and benign misinterpretations is to say that the former are characterized by failure to correctly identify a work’s genre.

441 Fish, “How To Recognize a Poem,” 324.
The greatest drawback to the genre-based approach to potentially destructive misinterpretation is that while it seems safe to say that correctly identifying a work’s genre is a necessary condition for proper interpretation, it is not a sufficient condition. Walton remarks that “category-relative interpretations do not allow aesthetic judgments to be mistaken often enough.” In other words, there are many flagrant misinterpretations of literary works that are not genre mistakes. For instance, consider a reader who reads and attempts to interpret *Ulysses*. The reader understands that *Ulysses* is a novel, that it is a modernist novel, and that it is a stream-of-consciousness novel. But for whatever reason, the reader badly misjudges the plot line of *Ulysses*. He believes that Stephen Dedalus and Molly Bloom murder Leopold Bloom. The stream-of-consciousness in the beginning of *Ulysses* is Stephen’s shortly before the murder, the stream-of-consciousness at the end of *Ulysses* is Molly’s shortly after the murder, and the long intermediate stream-of-consciousness that makes up the bulk of *Ulysses* is the hallucination Leopold has in the moments before he dies.

This interpretation of *Ulysses* is clearly false. It seems to be just as erroneous and potentially destructive as an interpretation of *Animal Farm* that takes it to be an odd children’s story or an interpretation of Fish’s reading list that takes it to be a poem. But no genre misidentification is involved in the case of *Ulysses*. In order to construe this misinterpretation as a genre mistake, the concept of genre would need to be recast in excessively and implausibly narrow terms. *Ulysses* would not just belong to the genre of modernist, stream-of-consciousness novel, but to something along the lines of “modernist, stream-of-consciousness novel, set over the course of one day, that begins with Stephen Dedalus’ stream-of-consciousness, spends most of the time following

---

Leopold Bloom’s stream-of-consciousness as he engages in various activities throughout the day, and ends with Molly Bloom’s stream-of-consciousness as she ends the day in bed, etc.” Meanwhile the misinterpretation of *Ulysses* would be the claim that the novel belongs to a genre along the lines of “modernist, stream-of-consciousness novel, set over the course of one day, that begins with Stephen Dedalus’ stream-of-consciousness before murdering Bloom, is subsequently taken up with Bloom’s lengthy hallucinations before dying, and ends with Molly Bloom’s stream-of-consciousness after the murder, etc.”

This attempt to construct extremely narrow genres calls to mind E.D. Hirsch’s intrinsic genre based criticism. According to Hirsch, the goal of interpretation is to discover the determinate meaning of every part of a given literary work. This is typically a lengthy process, and a reader cannot ascertain the determinate meaning of every part of a literary work until she has read the entire work. But in order to discover the work’s meaning, the reader must have some conception *before* she begins to read about what kind of or type of work it is.

The type or kind that a work belongs to includes its genre in the traditional sense; *Paradise Lost* is a “Christian humanist epic” for instance. Hirsch refers to this kind of genre as a work’s “extrinsic” or “heuristic” genre. If one is to properly interpret literary work *w*, one must know before reading what extrinsic genre *w* belongs to. But Hirsch does not consider knowledge of a work’s extrinsic genre enough; one also must know the intrinsic genre of *w* in order to properly interpret *w*. The intrinsic genre of *w* is still a conception of what type of work *w* is, but it is far more detailed than *w*’s

---

extrinsic genre and cannot be referred to by any simple label.⁴⁴⁶ For example, the extrinsic genre of *Paradise Lost* is “Christian humanist epic” but its intrinsic genre is “the particular type of Christian humanist epic” it is.⁴⁴⁷

While no reader can be aware of work w’s intrinsic genre before she begins to read and interpret w, Hirsch argues that she still must know w’s intrinsic genre before she finishes reading to the end of w if she is to have any hope of properly interpreting w. Hirsch believes one typically discovers a work’s intrinsic genre early on in the process of reading, as she is confronted by various features of the work which cause her to refine her conception of what type of work it is from the extrinsic genre conception she started off with. Once the intrinsic genre of a work has been correctly grasped, the reader can discover the determinate meaning of all subsequent parts of the work.⁴⁴⁸

If one accepts Hirsch’s take on intrinsic genre, misinterpretations such as the case where *Ulysses* is thought to describe the murder of Leopold Bloom can be seen as genre mistakes after all. While the extrinsic genre of *Ulysses* is simply modernist, stream-of-consciousness, novel, its intrinsic genre can be some long, detailed, label-defying concept like “modernist, stream-of-consciousness novel, set over the course of one day, that begins with Stephen Dedalus’ stream-of-consciousness, spends most of the time following Leopold Bloom’s stream-of-consciousness as he engages in various activities throughout the day, and ends with Molly Bloom’s stream-of-consciousness as she ends the day in bed, etc.” A reader who believes Stephen and Molly murder Leopold in *Ulysses* has indeed made a genre mistake, and has thus misinterpreted *Ulysses*.

⁴⁴⁶ Hirsch, *Validity in Interpretation*, 81-86.
⁴⁴⁷ Hirsch, *Validity in Interpretation*, 84.
While Hirsch is correct to note the importance of genre identification for guiding and shaping interpretation, his account of intrinsic genre is highly problematic. By redefining “genre” into convoluted messes like, “modernist, stream-of-consciousness novel, set over the course of one day, that begins with Stephen Dedalus’ stream-of-consciousness, spends most of the time following Leopold Bloom’s stream-of-consciousness as he engages in various activities throughout the day, and ends with Molly Bloom’s stream-of-consciousness as she ends the day in bed, etc,” proponents of intrinsic genre appear to be engaging in what is known as the “constructivist sulk.”

When confronted with an example of a misinterpretation that does not involve a genre mistake – *Ulysses* is about the murder of Leopold – the intrinsic genre proponent simply re-defines “genre” so that this misinterpretation and all other misinterpretations become genre mistakes. As a result, the very notion of genre collapses into meaninglessness.

5.5 Practical Consequences

Setting aside the difficult question of what characterizes potentially destructive misinterpretations, I would like to focus upon the practical, real-life consequences of artifactualism’s theory of interpretation. There is a nagging question that confronts any philosopher who completes a lengthy and narrowly focused project such as this – why should we care? In the case of this project, I believe we should care very much about artifactualism and its theory of interpretation. The destruction of literary works via

---

449 The constructivist sulk is also known as the “no true Scotsman” fallacy or the “every true Englishman” fallacy.
misinterpretation is not simply a hypothetical scenario used to illustrate some abstract philosophical point; it can and does indeed happen in real life.

It would be very difficult for a contemporary work of literature to be destroyed via misinterpretation. The dystopian scenario where *Animal Farm* no longer exists is very far-fetched and implausible. Such a scenario is so unlikely because it requires the irretrievable loss of so much information about the history of the Soviet Union. Every book, newspaper article, government record, archival document, personal account, and audio-visual resource that referred to the history of Stalinism and the Soviet Union would need to be thoroughly destroyed. This likely would entail the destruction of billions of pieces of information in various media. Given the vast number of sources of information about Stalinism and the ease at which information can now be preserved digitally and on the internet, it would be very difficult indeed for anyone – even a powerful oppressive government – to completely eradicate all knowledge of Stalinism.

While it is quite unlikely that contemporary works of literature will ever be destroyed via misinterpretation, misinterpretation is a genuine threat to the existence of many ancient literary works and works that originate from oral tradition. Consider the case of folklore. Folktales, even those that are transmitted solely in oral form, are considered literary works with the same dependencies that printed, published literary works have. In order for a folktale \( f \) to exist at time \( t \), at least one physical copy of \( f \)'s text (this “copy” can be a person who has memorized the tale) must exist at \( t \), and at least one person who can understand \( f \)'s language and properly interpret \( f \) must also exist at \( t \). A folktale requires proper interpretation just as much as any other literary work does. It is a mistake to believe, as one might be initially inclined, that all folktales are simple
children’s stories like “Little Red Riding Hood” that require little in the way of interpretation, unlike *Hamlet* or *Pride and Prejudice*. A good deal of folklore, particularly ancient folklore, is imbued with extremely complex meaning and significance. Note that works such as the *Odyssey* and *Genesis* are believed to have originated from oral tradition.

Sometimes the folklore of an ancient culture is relatively well preserved while other vital information about that culture has been lost to history. For example, a decent number of folktales from pre-Christian Slavic cultures have survived. The ancient Slavs had no system of writing, so these literary works survived via oral tradition until they began to be recorded in written form by ethnographers in the nineteenth and twentieth centuries. But because the ancient Slavs had no writing and rarely came into contact with cultures that did have writing, a great deal of important information about their culture, their beliefs, and their practices has been completely lost. There is a particularly severe dearth of information about the religious beliefs of the pre-Christian Slavs. Unlike with the ancient Greeks and Romans, we do not know most of the names of the Slavic pantheon. We do not know when two names refer to two separate gods or when one god has two names. We do not know when a name found in a folktale refers to a god or refers to something else. We do not know what gender some Slavic

---


deities were. We do not know what the attributes of the Slavic gods were or what their symbolism was.453

This creates a particular difficulty for interpreting some Slavic folktales. For instance, scholars believe that the eagle played some significant role in the religious beliefs of the ancient Slavs. It is not entirely clear, however, what or who the eagle was supposed to be. Some scholars believe the eagle was the warrior god (or a symbol of the warrior god). Other scholars believe the eagle was the sun god. Still others argue the eagle was the storm god. And yet other scholars claim the eagle was not a god at all but rather a malevolent nature spirit.454 Because there is so little information about the ancient Slavs’ religious beliefs it seems impossible for any scholar to know what the significance of the eagle in ancient Slavic belief was.

Now consider the Slavic folktale, “The Glass Mountain.” This story, which probably dates from the pre-Christian period, concerns an eagle that attacks a man who is trying to climb to a castle atop a glass mountain. The man kills the eagle and when the eagle’s blood spills on the ground, previously dead warriors come back to life.455

Is “The Glass Mountain” a religious story? Is the eagle in this story supposed to be an ancient Slavic god, and if so, which one? Or is the eagle simply an eagle with some magical properties? Given our lack of information about ancient Slavic religious beliefs, there is simply no way to answer these questions. This creates a serious ontological problem.

453 Ivanits, Russian Folk Belief, 12-15, 30-33.
454 Gimbutas, The Slavs, 164.
Assume that the eagle in “The Glass Mountain” is supposed to be the sun god. If this is the case, then it is appropriate to classify “The Glass Mountain” as a myth, a religious allegory, or simply as a story that contains a crucial allusion to a religious figure. Any proper interpretation of “The Glass Mountain” must recognize and acknowledge that the eagle is the sun god. This applies whether one believes proper interpretation is characterized by the acknowledgement of a work’s essential properties or whether it is characterized by the acknowledgement of a work’s genre. In the case of the former, being a religious allegory or containing a central allusion to a religious figure would seem to be among the essential properties of “The Glass Mountain.” In the case of the latter, plausible candidates for the genre of “The Glass Mountain” are myth and religious allegory.

Therefore, in order to properly interpret “The Glass Mountain,” a reader must acknowledge that the eagle is the sun god. A reader who fails to recognize this fact about “The Glass Mountain” will have provided a potentially destructive misinterpretation of the story. Of course, for all anyone knows, the eagle is not the Slavic sun god, but rather a symbol of the Slavic war god, much like Athena’s owl. This would change the meaning and significance of “The Glass Mountain” in a non-trivial manner. In this case, in order to properly interpret “The Glass Mountain,” a reader must acknowledge that the eagle is the symbol of the war god. And sometimes a magic eagle is just a magic eagle. If that is the case, then in order to properly interpret “The Glass Mountain,” a reader must recognize that the story is simply a fairy tale of sorts, meant for entertainment, not the communication of religious allegory. To interpret “The Glass Mountain” as a religious
allegory when it is, in fact, nothing of the sort would also be a potentially destructive misinterpretation of the story.

Whatever the proper interpretation of “The Glass Mountain” is, unfortunately no contemporary reader or scholar will be able to provide it. Too much of the relevant background information about the creation of this folktale and about the religious beliefs of the ancient Slavs has been irretrievably lost to history. If the story is a religious allegory, no contemporary reader or scholar can possibly know this. Likewise, if the story is a non-allegorical fairy tale, no contemporary reader or scholar can possibly know this. But both pieces of information – regardless of which one is actually the case – are absolutely essential for a proper interpretation of “The Glass Mountain.” And this has serious ontological consequences for “The Glass Mountain” itself.

In order for a literary work to exist at time $t$, at least one physical copy of the work must exist at $t$, and at least one person who can read the work’s language and properly interpret the work must exist at $t$. If the present moment, 2011, is to be $t$, then “The Glass Mountain” does not exist in 2011 and probably has not existed for a very long time. “The Glass Mountain” has become nonexistent not because no copies of its text exist – one is sitting on my bookshelf at this very moment. “The Glass Mountain” has become nonexistent not because no person can read its language – the tale has been orally transmitted in living languages like Polish and Russian for centuries and now exists in written form in other living languages like English. Rather “The Glass Mountain” no longer exists because no person can properly interpret it anymore. Every person who attempts to interpret “The Glass Mountain” will misinterpret it; a proper interpretation of this work requires background knowledge of Slavic religion that no living person can
possibly have. We are no more capable of correctly interpreting “The Glass Mountain” than readers in the Orwellian scenario can correctly interpret Animal Farm. And just as Animal Farm no longer exists in the Orwellian scenario, despite all the copies of its text written in living languages, “The Glass Mountain” no longer exists in our world, despite all the copies of its text written in living languages.

5.6 From the Golden Mountain to the Glass Mountain

Here we reach, at last, the conclusion of my project. My central claim has been that given the artifactualist account of the ontology of literary works, in certain circumstances the misinterpretation of a literary work can lead to the full-fledged ontological destruction of that work. Underlying this claim has been a lengthy philosophical journey through a series of seemingly unconnected issues in metaphysics, aesthetics, and literary theory.

In large part, this project is indebted to the work done in ontology by Brentano and Meinong over a century ago. As the first philosopher to seriously consider the ontology of all kinds of objects – real and unreal – Meinong has been a true catalyst for this project. Meinong’s work on the ontology of nonexistent objects such as the golden mountain opened the door for serious philosophical treatment of a closely related class of

---

456 I am assuming that a proper interpretation of a work is like knowledge; it is a justified, true, belief (or a set of justified, true, beliefs). Interpretations that happen to be correct but lack justification do not count as proper interpretations and cannot sustain literary works in existence. For example, imagine that “The Glass Mountain” is indeed a religious allegory involving the sun god. This information has been irretrievably lost to modern readers. But imagine that a modern reader makes a random guess about “The Glass Mountain,” that this story is a religious allegory about the sun god. Although this reader has a correct belief about the meaning and significance of “The Glass Mountain,” he has no justification for this belief. His belief does not qualify as a proper interpretation of “The Glass Mountain,” and does not enable “The Glass Mountain” to exist again in 2011.
entities – fictional characters. Whether one accepts Meinong’s arguments for realism about fictional characters or whether one prefers the simpler and more modern “linguistic case” for realism, there are sound reasons to believe that fictional characters like Hamlet and Sherlock Holmes are bona fide objects worthy of serious ontological consideration.

Although fictional realism owes its genesis to Meinong’s thought, we soon discover that Meinongian and neo-Meinongian theories of fictional characters are quite unsatisfactory. Meinongian theories of fictional characters, as well as the closely related Platonic theory of fiction, do not respect our ordinary intuitions and professional practices surrounding fictional characters. By placing fictional characters within the realm of the eternal and unchanging, Meinongianism conflicts with the deeply held intuition that fictional characters are contingent things created \textit{ex nihilo}, in time, by their authors.

A rival theory of fictional characters – the idealist theory – respects this intuition by claiming that fictional characters are ideal objects within the minds of their authors. Unfortunately the idealist account of fictional characters leads to counter-intuitive problems of its own. By placing fictional characters within the private realm of an author’s consciousness, idealism leads to the odd conclusion that no two readers ever read about the same character and that there are millions of Hamlets, millions of Sherlock Holmes, and so on.

Artifactualism emerges as a fine middle course between the Scylla of Meinongianism and the Charybdis of idealism. By claiming that fictional characters are contingent abstracta, or abstract artifacts, artifactualism respects both our intuitions that fictional characters are contingent upon the creative actions of their authors and our
intuitions that fictional characters are public objects that – once created – exist independently of the mental life of their authors.

As interesting as the artifactualist account of fictional characters is, it is only coherent within the artifactualist account of the ontology of literary works. According to Thomasson, a literary work depends upon its author to come into existence, but once it has been created, a literary work can continue to exist long after its author has died. Being a contingent object, a literary work still depends upon other entities to sustain it in existence. These entities, according to Thomasson, are a combination of spatio-temporal objects and certain mental states. Specifically, a literary work exists at time $t$ if and only if there also exist at $t$ at least one copy of the work’s text and at least one person who has the ability to read the language the work is written in and the “relevant background assumptions” required to understand the work.

The requirement that there be at least one reader with the proper background assumptions to understand a literary work is quite intriguing. Thomasson does not elaborate on this matter, but I argue that this condition is best understood as the claim that there must be at least one person who can properly interpret a literary work at a given moment in time if that work is to exist at that same moment in time. I develop Thomasson’s artifactualism a step further than she does. According to my revised version of artifactualism, in order for a literary work to exist at time $t$, at least one copy of the work’s text must exist at $t$ and at least one person who can read the work’s language and properly interpret the work also must exist at $t$.

My development of artifactualism leads to several interesting consequences. One consequence is that a literary work is dependent upon both its text and upon a proper
interpretation of it. This stands in sharp contrast to both textualism and constructivism. Textualism is the claim that a literary work is identical to its text, while constructivism is the claim that a literary work is identical to an interpretation (or class of interpretations) of it. Intuitive support for both textualism and constructivism can be found, but both theories entail unacceptably counter-intuitive metaphysical claims. Artifactualism’s account of the relationship between a work, its text, and its interpretations functions as an effective *via media* between textualism and constructivism. Artifactualism takes into account the intuitions supporting both theories while avoiding their negative consequences.

Another interesting consequence of the artifactualist theory of interpretation is that it leads to many literary works being temporally discontinuous or gappy. While gappiness initially appears to be a very serious problem for artifactualism – especially given artifactualism’s insistence upon the essentiality of a work’s origins – a closer consideration of the matter reveals the situation to be far rosier. Literary works can indeed be temporally gappy. Gappiness is not problematic for artifactualism given the unique way it distinguishes between literary works, interpretations, and the ways in which these respective entities are created.

My development of artifactualism also raises serious and difficult questions about what constitutes correct and incorrect interpretation. I grapple with this central question of the humanities, and while I do not presume to provide a conclusive answer to it, I do consider two approaches that seem relatively promising. At worst, these two approaches raise several interesting questions, especially regarding the possibility of individual essences for literary works and other closely related cultural entities.
But more interesting than that, I believe, is the practical import of artifactualism’s
theory of interpretation. The destruction of literary works via misinterpretation is not a
hypothetical example used to illustrate some arcane metaphysical point. It is something
that can and does indeed happen, “The Glass Mountain” being a plausible example of this
phenomenon. I strongly suspect that “The Glass Mountain” is not an isolated case; there
are many instances where nearly all information about the central beliefs and practices of
some human culture has been lost, with only the culture’s folklore surviving. In light of
artifactualism’s theory of interpretation, we would do well to reassess many of our
ordinary beliefs and professional practices surrounding ancient literature, folklore, and
other similar works.
BIBLIOGRAPHY


Lamarque, Peter, and Stein Haugom Olsen. *Introduction to Part IV of Aesthetics and the


-----. “New Historicism and Cultural Studies.” In *The Critical Tradition: Classic Texts


------. “Debates about the Ontology of Art: What are We Doing Here?” Philosophy Compass 1, no. 3 (2006): 245-255.


------. “Modal Normativism and the Method of Metaphysics.” Philosophical Topics (forthcoming).


------. “Quantification and Fictional Discourse.” In Empty Names, Fiction, and the


-----.


-----.


Curriculum Vitae

Allison Hepola

Universities Attended

August 1999 – May 2003  University of Notre Dame, B.A. in Philosophy
August 2003 – May 2011  Rutgers University, Ph.D. in Philosophy

Positions Held

August 2003 – May 2006  Graduate Fellowship, Rutgers University
July 2006-August 2006  Instructor – Summer Session, Rutgers University
August 2006 – May 2007  Teaching Assistantship, Rutgers University
December 2006 – January 2007  Instructor – Winter Session, Rutgers University
August 2007 – December 2007  Adjunct Instructor, St. Joseph’s University
December 2008 – January 2009  Instructor – Winter Session, Rutgers University
August 2009 – May 2010  Teaching Assistantship, Rutgers University
August 2010 – May 2011  Graduate Fellowship, Rutgers University