

FICTA AS MENTALIA:  
SURVEYING THEORIES OF FICTION  
IN SEARCH OF PLAUSIBLE ONTOLOGY

A Thesis

Submitted to the Graduate Faculty of the  
Louisiana State University and  
Agricultural and Mechanical College  
in partial fulfillment of the  
requirements for the degree of  
Master of Arts

in

The Department of Philosophy  
and Religious Studies

by  
Joseph Dartez  
B.A., University of Dallas, 2006  
August, 2009

## TABLE OF CONTENTS

ABSTRACT.....	iii
CHAPTER 1. INTRODUCTION.....	1
1.1 Realism, Anti-Realism, and Theories of Fiction .....	1
1.2 Ficta and Fictional Worlds.....	5
1.3 The Problem of Ontology and the Bifurcation in Speech.....	7
1.4 Fictional Incompleteness, the Sorites Paradox, and the Bifurcation .....	12
CHAPTER 2. AUGMENTED MODAL REALISM .....	16
2.1 Concrete Realism .....	16
2.2 Abstract Realism.....	22
2.3 The Failures and Successes of Augmented Modal Realism .....	29
CHAPTER 3. PRETENSE THEORY .....	33
3.1 Waltonian Pretense Theory.....	33
3.2 Fictionalist Pretense Theory .....	35
3.3 Objections to Pretense Theory .....	43
3.3.1 No Systematic Set of Principles of Generation.....	43
3.3.2 Pretense Theory Is Apparently Not Hermeneutic .....	44
3.3.3 No First-Person Authority Concerning Beliefs.....	45
3.3.4 Susceptibility to Empirical Defeat .....	46
3.3.5 Question-Begging Ontology .....	48
CHAPTER 4. COLLINGWOODIAN MENTALISM .....	54
4.1 Naïve Mentalism.....	54
4.2 Collingwood’s Theory of Art.....	59
4.2.1 Sensations and Emotions: The Constituents of Human Experience .....	60
4.2.2 Conscious Emotions: Expression <i>of</i> Emotion or Expression <i>as</i> Emotion?.....	62
4.2.3 Intellectual Experience” The Formalization of an Expressive Language.....	68
4.3 Applying Collingwood’s Theory .....	69
4.4 The Mentalist Interpretation .....	82
4.5 Conclusion .....	86
BIBLIOGRAPHY.....	92
VITA.....	94

## ABSTRACT

In the philosophy of fiction, there is a major debate between those who hold realist theories (theories which incur the existence of entities independent of human cognition) and those who hold anti-realist theories (theories which incur the existence of no independent entities). With this debate in mind, the primary goal of this essay is to construct a mentalist theory of fictional entities and worlds. Besides the mentalism constructed herein, three other theories are outlined and held to the same explanatory standards as the focal theory: two augmentations of modal realism (Lewis' concrete realism and van Inwagen's abstract realism), and a form of fictionalism which adopts as its fiction the abstract realist ontology. But these alternate theories are explored only insofar as they assist us in understanding which explanatory paths are fruitful and which not.

These four theories – concrete realism, abstract realism, fictionalist anti-realism, and mentalist anti-realism – are measured against each other using two standards: (a) their ability to explain the facts and intuitions which we have about fictional discourse, and (b) their ability to provide consistent interpretations of the literally contradictory sentences in fictional discourse. A theory which leaves one of these features unexplained or brings in *ad hoc* devices for the purpose of explaining facts about fictional discourse is considered inferior to a theory which does not. Concrete realism fails test (a) because it is a pure realism; abstract realism, on the other hand, passes test (a) because it incorporates a deflationary notion of fictional worlds and is thus not a pure realism but a hybrid of realism and anti-realism. Fictionalist anti-realism also passes both tests, but it fails to be a genuine anti-realism. These various failures and the solutions to these failures lead us to adopt a pretense-based Collingwoodian mentalism, which is a pure and genuine anti-realism. This realism is found to pass both tests, and its interpretations of sentences

spoken in fictional discourse are also found to be superior in both simplicity and faithfulness to their uninterpreted (and literally contradictory) counterparts.

## CHAPTER 1. INTRODUCTION

### 1.1 Realism, Anti-Realism, and Theories of Fiction

What sort of thing is Sherlock Holmes? In what kind of world does Dostoevsky's St. Petersburg reside? The short answer is a single word: "fictional"; but truisms provide little philosophical insight. In fact, an alternate subtitle to this essay could be "The Search for a Definition of 'Fiction'." For whatever reason, the inquiry concerning *fiction* (as opposed to the broader topic of art and aesthetics) did not garner much interest until the resurgence of metaphysics following the collapse of the project of logical positivism. In the context of this resurgence, only two positions concerning the ontology of fictional entities were taken seriously in Anglo-American philosophy, both of which entailed a belief in some set of mind-independent extant entities: that is, both of these positions are *realist* positions.

Before we say anymore about these realist theories of fiction, let us be precise about what we mean by realism. Jon Cogburn characterizes the broad features of realism neatly in his own matrix, under which realism about some discourse *D* is the following three claims:

- (1) Sentences in *D* are truth apt,
- (2) Some sentences in *D* are true, and
- (3) The truth or falsity of sentences in *D* are independent of human cognition.<sup>1</sup>

It is by virtue of (1) and (2) that (3) is typically taken to entail the independent existence of entities which guarantee the truth or falsity of sentences in *D*. For if these truth-values are fixed, then something must be "out there" to fix them.

Rejection of any of these three claims is sufficient for an *anti-realist* position concerning *D*. Rejection of (1) entails non-cognitivism; rejection of (2) entails error-theory, and rejection of (3) entails some form of relativism. In this essay, we will concern ourselves with breeds of anti-

---

<sup>1</sup> Personal communication.

realism which reject (3).<sup>2</sup> We are so concerned for two reasons. First, it is widely thought that sentences in the discourse of fiction are somehow true. To reject this intuition would be sufficiently counter-intuitive that such anti-realist theories should only be adopted when all other theories fail. Therefore, we assume that sentences in fictional discourse are both truth-apt and sometimes true, and we will see that this assumption has often driven the debate about realism and anti-realism in fiction. Second, the three major theories considered herein, as well as the fourth one constructed, all assume both (1) and (2). So for the purposes of this essay, we simplify the defining claims of realist positions and anti-realist positions to the acceptance of (3) and the rejection (3), respectively.

This simplification of the debate between realism and anti-realism entails that our concern is *ontology*: it is the *entities* to which realists about fiction are attached which will constitute the subject of our discussion. So the patented realist move of positing entities which

---

<sup>2</sup> We do not go so far as Crispin Wright has gone in setting the limits of “realism” (*Truth and Objectivity*, [Cambridge, Mass. & London: Harvard University Press, 1992]. Hereafter cited in text as *TO*). On Wright’s account, a discourse is minimally truth-apt if it adopts the platitude that facts somehow correspond to reality. And, naturally, adoption of the platitude does not entail acceptance of the platitude. In short, minimal truth-aptness is the deflationary theory of truth – a distinctly anti-realist position. But the Best Explanation Constraint is a method of expanding this minimal concept of truth-aptness into a form of realism:

A discourse is more than minimally truth-apt only if mention must be made of the states of affairs which it concerns in any best explanation of those of our beliefs expressed within it which are true (*TO*, 177).

In short, the BEC allows us to gauge the influence that our discourses have on our beliefs about those discourses and about the rest of the world. The BEC is the litmus test for the real influence that our discourses have upon us, and much of modern science has passed this test with flying colors. It is the BEC which often brings us to be realists about protons and electrons – for they explain the workings of electronic devices.

For a discourse to pass the BEC, it must have some degree of width of “cosmological role” – an influence in the real world. While the BEC is, for Wright, nothing more than an intuition, this intuition is relevant *because* it points to the width of a discourse’s cosmological role. A discourse with a wide cosmological role will typically explain some or all of the following: (a) cognitive effects, (b) pre-cognitive sensuous effects, (c) effects on physically interactive agents, and (d) brute effects on inanimate matter. These are the marks of a realist discourse, for Wright. But our concept of realism is not intended to be so nuanced, for it is a simpler concept. In fact, our initial concept of realism is that creature which naively grazes while it is centered in the crosshairs of *Truth and Objectivity*: Platonism (realism) versus nominalism (anti-realism). But just as Wright concludes that both sides have made unfortunate leaps beyond their own epistemic capacities, so we will conclude that both sides have tended to make existence claims which are too extreme – a tendency which will make an ontology both cumbersome and counter-intuitive. This is not to say that Crispin Wright has failed to end this debate; rather, if Wright has ended the debate, then theorists of fiction have not yet seen it happen. Therefore, one of the goals of this essay is to bring theorists of fiction face to face with the end of this debate.

secure the independent truth or falsity of propositions about fiction is the very move that we will take anti-realists about fiction to be concerned with blocking. We begin by taking as our body of evidence concerning fictional discourse the set of all sentences spoken about fictional characters and worlds. Of these sentences, we take as relevant the ones which are spoken as if they are true. This set of sentences is ultimately to be the evidence by which any theory of fiction must be tested. If a theory should account for all of the evidence, then we will expect a definition of the context in which these sentences are spoken – the so-called “fictional” context. We do not assume a prior knowledge of what it means for a sentence to be spoken in a “fictional” context, so any demand for such an account amounts to a demand to favor one theory over another pre-theoretically. Hence, theories of fiction are often characterized by the ontological or semantical structure which is posited as an explanation for sentences about fiction and the ways in which these sentences are either true or false.

This structure can fail in numerous ways. The most drastic way in which it can fail is by rendering false a class of sentences within fictional discourse which are spoken as if they are true. Thus, theories of fiction are often measured by the ways that they must interpret these sentences such that they come out true. Another way that a theory of fiction can fail is by ignoring a crucial intuition concerning fictional discourse. This is considered failure because counterintuitive theory is both more cumbersome to deploy and less likely to be believed than an intuitive one. Another way a theory of fiction can fail is that it might deny observable facts concerning fictional discourse. And a final, more traditional, way that a theory of fiction might fail is by entailing contradiction. These potential failures will all be relevant to our essay.

The primary goal of this essay is to develop an anti-realist theory of ficta and fictional worlds which is internally consistent and fails in none of the ways mentioned above. Again, in

the name of preserving the intuitions we have about the veracity of sentences spoken in fictional discourse, we take anti-realism concerning ficta to be a matter of independence from or dependence upon human cognition. To put this concept modally: realism about ficta and fictional worlds entails that:

- (4) It is possible that there is a time  $t$ , such that there exists at least one fictum at  $t$ , and there exists no human minds at  $t$ .

Let us, for the purposes of this essay, grant this assumption that anti-realism about ficta amounts to relativism, for it is indeed an assumption, as there is no reason to think that functional non-cognitivist theories or error theories about fictional discourse, despite their natural counter-intuitiveness. In an effort to meet this goal of a relativistic theory of ficta and fictional worlds, we take the most intuitive route: the position that ficta and fictional worlds exist in the mind as thoughts or *mentalia*. Thus, we may refine the statement of our primary goal: to develop a mentalist theory of ficta and fictional worlds which is internally consistent and fails in none of the ways mentioned above.

In the effort of meeting these goals, we will examine the alternate theories as much as is needed and we will borrow from them as necessary. Within this essay, we will use two augmented modal realist theories of fiction (concrete realism and abstract realism) as meta-theoretical evidence for our choice to pursue pretense theory as a model upon which to build a fictional mentalism. But pretense theory comes prepackaged in the Tradition with its own theory of fiction, so we will first attempt to fit this pretense theory out with a mentalist ontology. We will discover that this naïve pretense mentalism fails to meet our requirements for a successful theory, for it is susceptible to the mentalist's most troubling objection: there is no explanation for agreement in fictional discourse which is not *ad hoc*. Therefore, we will pursue a sophisticated mentalism which borrows heavily from R. G. Collingwood's theory of art.



## 1.2 Ficta and Fictional Worlds

To what do fictional names refer? It is widely thought by non-philosophers that proper names refer to the existing objects which bear those names and if exception is made to this rule it is for the sake of names in fictional discourse. Indeed, what else could a name do but designate an object? Catering to the intuition that fictional discourse is parasitic upon normal discourse somehow – for how could a fictional world even be conceived without the prior experience of the actual world? – let us say that the standard use to which a name is put is the designation of an existing object, where existence is construed as realist in the sense proposed above. Of course, this still does not rule out the possibility that fictional names *do not*, in fact, refer to any existing object. This concern, however, we will not directly address here.

Among fictional characters, there are some which are entirely fictitious – such as Sherlock Holmes – and some which are parasitically fictitious – such as Shakespeare’s Julius Caesar. *Julius Caesar* is not intended to be a historical document, so there is a legitimate sense in which the Shakespeare’s Caesar is independent of the historical Caesar: Shakespeare has the ability to send Caesar to outer space if he so chooses, and it matters not that the historical Caesar never left Earth. Is it the case, then, that the name ‘Caesar’ refers to the historical Caesar? Surely not. If it did, then many fictional sentence about Caesar would be false. But fiction is more subtle than that: we distinguish between the sentences which are true about the historical Caesar, the sentences which are true about Shakespeare’s Caesar, and the sentences which are true about both. But we need some way to distinguish the actual Julius Caesar from the fictional Julius Caesar, so that we may allow sentences about each to be true: we must separate the fictional and historical discourses. The theoretical structure most suited to this task is the concept of fictional worlds. So before we address the problem of the ontology of ficta, we must first describe the

context in which we find such entities – we must say something about the worlds which separate ficta from each other and from actually existing entities (hereafter: *actualia*).

Like any definition of art in modern aesthetics, an adequate definition of fiction assumes the prior existence of a *theory* of fiction. Therefore, we will attempt to define fiction later and for now we will be content to focus on uncontroversial canonical examples of fiction such as Dostoevsky's *Crime and Punishment*, Conan Doyle's stories about Sherlock Holmes, and Shakespeare's *Julius Caesar*. *Prima facie*, works of fiction take the form of a description of the doings and happenings of people or creatures in a world, but this description is a *false* description – these doings and happenings were *not* done and did *not* happen in the actual world. Rather, they take place in a *fictional world*, where the entities to which these happenings occur reside – and these worlds may differ from each other as much as they differ from the actual world. The world in which Sherlock Holmes lives is a world in which it is (in a non-technical way) *practically possible* for a seasoned detective to solve every case he has ever taken on (this is a practical impossibility in the actual world) – and this is one of the rules of Sherlock Holmes' world which circumscribes the potential events of the world. Conan Doyle expanded the practical rules of the actual world when he invented his Sherlock Holmes world, for in Conan Doyle's Sherlock Holmes world there exists a man of such genius that no mystery can elude his acumen.

So it is apparent that ficta are subordinate to the rules of the worlds in which they reside. Therefore, the concept of fictional worlds must be explored with as much fervor as the concept of fictional characters, as it appears that our concept of ficta will piggyback upon our concept of fictional worlds. To this end, it will behoove us to tentatively appropriate Kendall Walton's concept of fictional worlds:

Each fictional world is associated with a particular class or cluster of propositions – those propositions that are fictional in that world. Some will be tempted to identify fictional worlds with these clusters of propositions.... This would make fictional worlds look very much like possible worlds, for a standard way of construing the worlds is as sets of propositions. But fictional worlds are not possible worlds. Two differences, especially, have been discussed elsewhere: Fictional worlds are sometimes impossible and usually incomplete, whereas possible worlds (as normally construed) are necessarily both possible and complete.<sup>3</sup>

The cluster of propositions about which Walton speaks is not the group of sentences which constitute the text; rather, it is the set of propositions which are entailed *by* the (fictionally true) sentences in the text. The way in which these propositions are entailed by the sentences in the text is one that must be clarified in any theory which adopts Walton's notion of fictional worlds, but because more than one theory of fiction will be outlined, we will stave this task off until we have better clarified our theory. Let us merely say that while all propositions which constitute a fictional world are somehow *true*, not all sentences in the text will be true (e.g. when a character speaks a lie). As Walton observes in the passage above, one defensible position *identifies* the cluster of propositions with the fictional world, but this need not be the case.

In addition to the association with a cluster of propositions, Walton's notion of *incompleteness* will prove to be critical to our discussion of the ontology of ficta and fictional worlds.

### **1.3 The Problem of Ontology and the Bifurcation in Speech**

We can immediately classify two types of positions concerning ficta and fictional worlds: realism and anti-realism. In accord with the sketch above, realists believe that ficta and fictional worlds are entities that exist apart from or beyond mere human cognition. Conversely, anti-realists believe that whatever the makeup of the world external to the human mind, there are certainly no fictional characters or worlds to be found in it. So, one might ask, why be a realist?

---

<sup>3</sup> Kendall Walton, *Mimesis as Make-Believe*, (London, England: Harvard, 1990), p. 64. Hereafter cited in text as *Mimesis*.

In order to answer this question, we must first lay out the ontological problem as it is traditionally understood.

There are generally thought to be three classes of sentences involving ficta and fictional worlds.<sup>4</sup> Though any theory which adopts Walton's model of fictional worlds will need to determine how propositions are entailed by the sentences of a text, let it be noted that we take the example sentences offered herein to be uncontroversial examples of sentences which entail propositions. And because all of the propositions constituting a fictional world are true, these sentences must all be somehow true.

The first class, the so-called "fictional" sentences, are just the sort that are found in works of fiction: they describe what is the case *from within* that fictional world. Thus, the defining feature of these sentences is that they describe a fictional world as if it were the real world. Whether these sentences are actually found in texts is not relevant to whether they are fictional sentences, though there is certainly some priority given to the particular sentences which constitute a complete fictional text (i.e. the sentences that Dostoevsky eventually published as *Crime and Punishment*). Some examples of this first class are:

- (F<sub>1</sub>) (a) Rodion Raskolnikov is a murderer.
- (b) Sherlock Holmes plays the violin.

What is most curious about these sentences – and what will later prove to be most troublesome – is that they are spoken in precisely the same way as sentences which describe actualia. So there appears to be a sense in which ficta are the *same* as actualia. This similarity in usage is the defining feature of the fictional class of sentences about ficta.

---

<sup>4</sup> Though these groupings are not unique to Stuart Brock's work, their designations have been borrowed from his essay "Fictionalism about Fictional Characters," *Noûs* 36, 1 (2002): 1-21. Hereafter cited in text as "Fictionalism".

The second class of sentences about ficta and fictional worlds is sometimes called the “critical” class of sentences. These sentences are used by persons in the actual world who refer to ficta and fictional worlds in some critical way. In other words, these sentences are spoken with an awareness that the speaker is *outside* the fictional world. They might be used to compare ficta and fictional worlds to the actual world (inter-comparative critical sentences), to another fictional world (intra-comparative critical sentences), or they might be used non-comparatively to analyze the structure of a fictional world (analytic critical sentences). The defining feature of comparative critical sentences is that they make reference to ficta in a *cross-world* fashion (regardless of whether the world is fictional or actual). On the other hand, the defining feature of analytic critical sentences is that they could *not* have been spoken about the real world: fictional characters are recognized as characters and not as persons. Some examples of critical sentences are the following:

- (F<sub>2</sub>) (a) Sherlock Holmes is a more clever detective than Porfiry Petrovich.
- (b) Sherlock Holmes is more famous than any real detective.
- (c) Rodion Raskolnikov is an archetypal character.

The third class of sentences about ficta and fictional worlds is the existential class. These sentences are typically negative atomic propositions asserting that ficta and their worlds do not exist. Normally, such sentences would be uttered casually and pre-philosophically. For example,

- (F<sub>3</sub>) (a) Rodion Raskolnikov does not exist.
- (b) Sherlock Holmes does not exist.

All three of the foregoing classes of sentences are instances of sentences which most of us who are familiar enough with the texts would assent to without a second thought. The class of sentences that stand out most prominently for the debate between realism and anti-realism in fiction is the third class, the existential class. ‘Sherlock Holmes does not exist’ is a literal existence claim which seems to end the debate immediately – in the anti-realist’s favor.

However, should we be willing to grant literal truth to this claim, we will be required to determine how it can be the case that sentences of the first and second classes can still be true. If it we allow fictional names to refer, then Raskolnikov and Holmes *do* exist, and thus we would have direct contradiction with our third class – for how could Raskolnikov be a murderer if he is nothing at all? A secondary problem is that we would also be inclined to think that non-existent things do not have properties typically attributed to actualia (such as being-a-murderer). These kinds of properties are had by physical objects in the actual world, which does not tell us much about what we mean when we say that Raskolnikov is a murderer. Nevertheless, it is certain that (F<sub>1</sub>a) must be literally false if (F<sub>3</sub>a) is to be literally true. The problem of the ontology of fiction, then, is this: a simultaneous literal translation of all three types of sentences will inevitably lead us to contradiction. Hence, the appropriate interpretation of these three classes of sentences has been a central feature of the debate about the ontology of ficta and fictional worlds.

But what constitutes an “appropriate” interpretation of a class of sentences about fiction? What will be our standards of appropriateness? We have already hinted at the answer to this question in our descriptions of the fictional and critical classes of sentences. Aside from the fact that we make negative existential claims about ficta, we also tend to describe them in two very distinct ways: we speak of ficta sometimes *as if from within* a fictional world and sometimes *as if from without*.

When one reads *Crime and Punishment*, one is left with the impression that if Dostoevsky were asked what color Raskolnikov’s hair is, he would have an answer. This phenomenon suggests that fictional worlds are to be thought of as fully fleshed out and complete when described from within. There is nothing, in principle, that prevents Dostoevsky from describing every single detail of the fictional Petersburg in *Crime and Punishment*, except the

infinite nature of the demand. This in lack of in principle limit to the determinacy of a fictional world is based on the fact that an author cannot be wrong when he writes fictional sentences. If we take this apparent in-principle determinacy as a fact, then this fact suggests the intuition that the details of fictional worlds seem to be “out there” *in* the fiction world, regardless of whether they are written by Dostoevsky or not. And even if they are not “out there”, at least they are to be thought of as if they are “out there”, for an in-principle determinate world may be, in fact, more determinable than its canonical description demands. In short, when we speak fictional sentences as if from within the fictional world, we seem to be participating in the fiction, which entails that we have the freedom to imagine the features of this world to an unlimited degree of determination. Thus, when we speak fictionally, or as if from within, we assume the position that the fictional world is *complete and fully determined*.

On the other hand, when a literary critic speaks about *Crime and Punishment*, provided she is a faithful critic, she will never assume any proposition concerning the fictional world which is not at least implied by Dostoevsky in the sentences of the text. She will never assume that she knows how many hairs are on Raskolnikov’s head, because no mention is made of this number in the text itself. Thus, when we speak critically, or from without, we assume the position that fictional world is *incomplete and underdetermined*. To this extent, the way we speak about ficta is very distinct from the way we speak about actualia. Let us call these opposite positions concerning the completeness and determination of fictional worlds the *Bifurcation in Speech*.

The Problem of the Ontology of Ficta – viz. that we do not seem to speak consistently about *whether* ficta exist or about *how* they exist if they do – must be solved with the Bifurcation in mind. Not only must our ontology be sufficiently robust to provide an interpretation under

which all three classes of sentences can be simultaneously true, but *it must also be sufficiently robust to explain the fact of the Bifurcation in Speech*.

#### **1.4 Fictional Incompleteness, the Sorites Paradox, and the Bifurcation**

Roy Sorensen has developed an argument that bolsters our choice of desiderata for a theory of fiction by demonstrating that the Bifurcation is more than a mere intuitive illusion.<sup>5</sup>

This argument shows that a good theory of fiction must account for the similarity between sorites paradoxes produced by fiction and standard sorites paradoxes. Sorensen begins with an example of a traditional sorites argument: Assume that:

(5) 1 minute after noon is noonish.

Assume also that

(6a) If 1 minute after noon is noonish, then 2 minutes after noon is noonish.

Thus, we may construct the rule

(6) If  $n$  minutes after noon is noonish, then  $n + 1$  minutes after noon is noonish.<sup>6</sup>

We may now iterate as many times as necessary to arrive at the absurdity that

(7) Six hundred minutes after noon is noonish.

Yet, six hundred minutes after noon is *not* noonish, because ten o'clock PM is well into the night. Hence, the sorites argument produces a paradox.

Sorensen observes that a sorites argument can be constructed based on fictional incompleteness. Let us assume that Sherlock Holmes is determinately complete in his own fictional world. We begin with a safe statement such as

---

<sup>5</sup> Roy Sorensen, "Fictional Incompleteness as Vagueness," *Erkenntnis*, 34 (1991): 54-72. Hereafter cited in text as "Incompleteness"

<sup>6</sup> Sorensen makes sure to point out that "the negation of the induction step is equivalent to the assertion that there is a precise division point between times that are noonish and times that are not" (56ff). But, of course, imprecision is essential to the concept of 'noonish'. Thus, we cannot dismiss this argument on account of a false induction in (6). He responds similarly to the induction step in the fiction sorites ("Incompleteness" 57).



(8) Sherlock Holmes had at least 0 hairs on his head.

Because hairs cannot come in a quantity fewer than 0, this statement is true by definition. But the number of hairs on Sherlock Holmes' head is not specified in any story, and it is likely that he had at least one hair, because nearly all human beings do:

(9a) If Sherlock Holmes had at least 0 hairs on his, then he had at least 1 hair on his head.

And we can assume with nearly equal confidence that:

(9b) If Sherlock Holmes had at least 1 hair on his head, then he had at least 2 hairs on his head.

This step can be iterated as many times as necessary, so we reformulate our second premise:

(9) If Sherlock Holmes had at least  $n$  hairs on his head, then he had  $n + 1$  hairs on his head.

But now we will find ourselves knee deep in absurdities such as

(10) Sherlock Holmes had at least one billion hairs on his head ("Incompleteness" 56).

Because Conan Doyle never mentioned how many hairs were on Sherlock Holmes' head, there is no definite "least number" of hairs. Nevertheless, no human being has more than a few hundred thousand hairs on her head. Again, we have paradox.

The diagnosis of standard sorites arguments is that they are caused by vagueness, and the obvious similarity between the two arguments suggests that vagueness is also involved in the fiction sorites argument. According to Sorensen, "vagueness theorists have not displayed much sympathy for the view that objects can be vague. Generally, they take vagueness to be a feature of predicates rather than a feature of the things to which predicates apply" ("Incompleteness" 56).<sup>7</sup> Should a fiction theorist attempt to *explain* the fiction sorites via a concept of vagueness,

---

<sup>7</sup> I quote: "Despite general suspicion of the notion of vague objects among vagueness theorists, detailed discussion of the notion has centered around a single argument advanced by Gareth Evans: 1978, 'Can There be Vague Objects?', *Analysis* 38. Although Evans' argument has been criticized as inconclusive, few have gone on to assert

such a fiction theorist will be guilty of assuming that an inductive fact can explain its own existence. For “vagueness is the diagnosis of the sorites paradox, not a defining condition of it” (“Incompleteness” 58). Therefore, it is the *vagueness itself* – that paradox-producing phenomenon – that must be explained by a theory of fiction. So any theory of fiction whose treatment of the fiction sorites does *not* diagnose the paradox as a consequence of vagueness (like all sorites arguments) must not only explain why the two paradoxes are to be treated differently (on pain of an *ad hoc* theory), but it must also tell us from whence the vagueness comes. For Sorensen, the natural response is to simply accept that the two are both a consequence of *semantic* vagueness.

Sorensen concludes that “all fictional incompleteness is vagueness.”<sup>8</sup> This interpretation of fictional incompleteness suggests that the fictionally *incomplete* character, Sherlock Holmes, is really just a *complete* character *vaguely* portrayed. In other words, when we speak of Sherlock Holmes from within his fictional world, he has a definite number of hairs on his head, just as he has a definite number of freckles on his right hand. There is no vagueness in the real world, only vagueness in our *talk* about the real world. Similarly, there is no vagueness within a fictional world – only vague description of that fictional world. When we speak of Sherlock Holmes critically, comparing him to other characters and to the real world, it seems that he is something like an incomplete bundle of properties which are typically exemplified by humans.

---

that there are vague objects. One member of this small group is Bertil Rolf who discusses the issue in his *Topics on Vagueness* (“Incompleteness”, 71, endnote 12).

<sup>8</sup> Sorensen’s summarizes his argument as follows: “All fictional incompleteness is a matter of some questions about a story being left indeterminate. All such indeterminacy is (either directly or indirectly) numerical. Hence all fictional incompleteness implies numerical indeterminacy. Wherever we have numerical indeterminacy, it is possible to construct a fiction sorites argument. So all fictional incompleteness breeds fiction sorites arguments. But all fiction sorites arguments arise because the indeterminacy is a form of hidden vagueness. Therefore, we conclude that all fictional incompleteness is vagueness” (“Incompleteness” 61).

The distinction between speaking of a character from within a fictional world and speaking of a character from without, a distinction rejected by Fred Adams, Gary Fuller and Robert Stecker as needlessly disunified,<sup>9</sup> is actually a structure which is capable of bearing the type of vagueness which we find in the distinction between objects and predicates. Fiction sorites paradoxes demonstrate that the within/without Bifurcation that seems to occur across our three classes of sentences is a feature that we are not justified in rejecting. In fact, it is apparent that the Bifurcation is much more than an intuition, it is, as we suggested earlier, an *empirical fact* about fiction which will need to be explained by any serious theory of fiction.<sup>10</sup> What is required of such an explanation is that it addresses both sides of the Bifurcation. On the one side, the *disengaged* side, we have incompleteness, now characterized as vagueness, which allows us to construct fiction sorites arguments. On the other side, the *engaged* side, we have the objects *about which* our speech is vague. When we speak about ficta from this other side, our speech is *only* subject to *standard* sorites arguments; fiction sorites arguments do not apply. For, as we have said, Watson could simply count the hairs on Sherlock Holmes' head. And if Evans' argument is to be trusted, these objects are not themselves vague.

---

<sup>9</sup> "The Semantics of Fictional Names," *Pacific Philosophical Quarterly*, 78 (1997): 128-148. Hereafter cited in text as "Semantics".

<sup>10</sup> It bears noting that Braun maintains a distinction between Holmes<sub>1</sub>, a non-referring name, and Holmes<sub>2</sub>, a referring name. On Braun's account, Holmes<sub>2</sub> refers to an abstractly existing fictional character ("Empty Names..." 609). While this distinction does explain the Bifurcation, it also explicitly relies on a realist ontology, so Braun's explanation of the bifurcation is quite useless to the anti-realist.

## CHAPTER 2. AUGMENTED MODAL REALISM

### 2.1 Concrete Realism

Sorensen's reflections on the completeness of fictional worlds leads us to suppose that our critical sentences about fictional worlds are a consequence of the vague description of a genuinely complete world filled with determinate objects which lack any vagueness. The simplest theory which can capture this intuition is concrete realism. For the concrete realist, fictions are concrete entities which populate other concrete worlds just like ours, but to which we have no physical access. In this sense, and in many others, concrete *fictional* realism is strongly related to concrete *modal* realism. Indeed, the same man defends both positions: David Lewis.

For Lewis, although it is possible that a man in the actual world might bear all the properties of Sherlock Holmes, this man is *not* Sherlock Holmes, because he lives in the wrong world. Conan Doyle imagined a *fictional world* in which Sherlock Holmes lived, so the speech act 'Holmes' is only an ordinary proper name in the world in which he lives. When the name "Holmes" is written in a context outside his home world, it only refers to that concrete being which fills the function of Holmes in that world:

The sense of "Sherlock Holmes" as we use it is such that, for any world  $w$  where the Holmes stories are told as known fact rather than fiction, the name denotes at  $w$  whichever inhabitant of  $w$  it is who there plays the role of Holmes. Part of that role, of course, is to bear the ordinary proper name "Sherlock Holmes". But that only goes to show that "Sherlock Holmes" is used at  $w$  as an ordinary proper name, not that it is so used here.<sup>11</sup>

Lewis believes that Sherlock Holmes is essentially a *person*, but he is a person whom we think about in *functional* terms. Lewis explains the incompleteness of fiction as a simple matter of the underdetermination of fictional worlds. While there are a finite number of Holmes stories, their collective description of the fictional world of Sherlock Holmes is not sufficiently

---

<sup>11</sup> David Lewis, "Truth in Fiction," *American Philosophical Quarterly*, 15, 1 (1978): 37-46. Hereafter cited in text as "Truth in Fiction".

comprehensive to determine all the features of a fully fleshed out concrete reality. There are many descriptions of places and events whose details have been left unmentioned. Thus, there are potentially infinitely many *determinate* Sherlock Holmes worlds – corresponding to all the various elliptical descriptions in Conan Doyle’s writing – and the stories describe *all of them* (“Truth in Fiction” 39). This entails that there are numerous concretely existing Sherlock Holmeses about whom all the stories are true. Some Holmeses have fewer hairs than others, some Holmeses have fewer 5<sup>th</sup> cousins than others, and so on. The only propositions that can be said to be true of *all* Sherlock Holmeses are the propositions expressed in Conan Doyle’s stories. Succinctly,

A sentence of form “In fiction  $f$ ,  $\phi$ ” is true iff  $\phi$  is true at every world where  $f$  is told as known fact rather than fiction (“Truth in Fiction” 41).

For Lewis, then, propositions are associated with fictional sentences in just the same way as propositions are associated with any other kind of sentence in a natural language. We will assume that this means of association is uncontroversial and continue.

Let us consider the concrete realist’s interpretations of our three classes of sentences about ficta and fictional worlds. A concrete realist paraphrases existential sentences thus:

- (F<sub>3</sub>) (a’) Rodion Raskolnikov does not *actually* exist.  
(b’) Sherlock Holmes does not *actually* exist.

In this paraphrase, ‘actual’ simply means ‘in *this* concrete world’.

Concrete realists may interpret fictional sentences almost literally, because fictional sentences are taken as describing a fictional world from within: they are the told as fact, just not in *this* world. On Lewis’ account, Conan Doyle imagined an existing determinate world which he described in his stories, and in this determinate world, ‘Sherlock Holmes’ is a proper name that refers to a detective. More specifically, Conan Doyle pretended to be telling the Sherlock

Holmes stories as known fact – he pretended to be Watson. So the propositions which are associated with the fictional sentences are all true, but they are world specific, so our interpretation is minimal. Let  $w$  designate the world(s) where the fiction is told as known fact, and let this fiction be designated by a subscript:

- (F<sub>1</sub>) (a') In  $w_{CP}$ , Rodion Raskolnikov is a murderer.  
 (b') In  $w_{SH}$ , Sherlock Holmes plays the violin.

Critical sentences must be interpreted in function terms, for Lewis. We will therefore designate proper names which refer functionally with quotation marks. A further convenience for concrete realism is that properties are not world-indexed, so cross-world comparison poses no major problem:

- (F<sub>2</sub>) (a') In  $w_{SH}$ , 'Sherlock Holmes' functions as a more clever detective than 'Porfiry Petrovich' does in  $w_{CP}$ .  
 (b') In  $w_{SH}$  'Sherlock Holmes' functions as a more famous detective than any detective does in the actual world.  
 (c') In  $w_{CP}$  'Rodion Raskolnikov' plays an archetypal role.

I should stress that our subscripts, which appear to isolate a *single* world do not, in fact, have this function.  $w_{CP}$  refers to an entire set of worlds, all different from each other, but each of which fits the description in *Crime and Punishment*. It is for this very reason that the concrete realist has no choice but to assume that properties remain constant across worlds. So, to summarize, we have:

### Interpretation under Concrete Realism

- (F<sub>1</sub>) (a') In  $w_{CP}$ , Rodion Raskolnikov is a murderer.  
 (b') In  $w_{SH}$ , Sherlock Holmes plays the violin.  
 (F<sub>2</sub>) (a') 'Sherlock Holmes' functions as a more clever detective in  $w_{SH}$  than 'Porfiry Petrovich' does in  $w_{CP}$ .  
 (b') In  $w_{SH}$  'Sherlock Holmes' functions as a more famous than any detective does in the actual world.  
 (c') In  $w_{CP}$  'Rodion Raskolnikov' plays an archetypal role.  
 (F<sub>3</sub>) (a') Rodion Raskolnikov does not *actually* exist.  
 (b') Sherlock Holmes does not *actually* exist.

Concrete realism has two distinct and strongly attractive benefits. The first is that, *prima facie*, it neatly accommodates the intuition that fictional worlds are. The second is that in the concrete realist's interpretations, *all* proper names refer, so there is no need to trouble with the consequences of empty names.

In addition to these benefits, concrete realism is capable of explaining the disengaged side of the Bifurcation quite smoothly. Because of the limitations of human expression, no author can pinpoint a fictional world down to the smallest detail – this would take an eternity to do. Thus, the perceived incompleteness is actually a form of vagueness (just as Sorensen demands): the author simply did not determine the fiction precisely enough to narrow the referent worlds down to unity. Nevertheless, the author still writes his story as if he were in a single (fictional) world – and he *must* do so, for such are the limitations of humankind.

However, concrete realism falters when it attempts to address the engaged side of the Bifurcation. Lewis describes Arthur Conan Doyle as having *pretended* to be Watson describing the doings of his detective roommate. But what this really amounts to in Lewis' theory is that Conan Doyle pretended that the world he described was complete, or at least complete in principle. But even in fictional sentences,  $w_{SH}$  does not pick out a *single* world; it picks out a set of worlds which fit the functional description. Lewis can avoid this worry by privileging one Sherlock Holmes world over the others and designating this as the world that Conan Doyle pretended to be in when he pretended to be Watson. But this move is not very promising. This arbitrary privileging of one world over another does disservice to Lewis' explanation of the disengaged side of the Bifurcation. If a single world is privileged, then we have no need of functional description. We can simply refer to Sherlock Holmes the man, because we have a single privileged world (though we do not know which) identified by Conan Doyle. Until Lewis

provides an account of what it could mean to pretend to be in a fictional world (singular), it appears that we will not be able to reconcile this concept of pretense with Lewis' identification of  $w_{SH}$  with a *set* of fictional worlds. Either *Crime and Punishment* picks out a single determinate world or it picks out a set of determinate worlds, so Lewis can only accommodate one side of the Bifurcation.

Moreover, concrete realism fails to explain one of the most central intuitions about fictional discourse: the notion that we create or generate fictions and fictional worlds is entirely unsupported under this theory. It is already absurd enough that we are positing an infinity of worlds just like ours, so if Lewis asks us to believe that we also *create* these worlds, then we should not be faulted if we respond "no, thank you." But he won't ask this of us. If we can generate concrete fictional worlds by imagining them, then other worlds could generate ours by imagining it, and Lewis must have recognized this absurdity as such. Lewis is hesitant to privilege fictional (and possible) worlds beyond the mere happenstance that we live in the "actual" world. Other worlds are populated by beings which are just as real as we are, so, as with all other features of concrete realism, if any powers are granted to us in this world, they must also be granted to our counterparts in other worlds. Consequently, we cannot posit the ability to create fictional worlds.

This is not the only potential absurdity with which concrete realism is faced. Until now, I have been quite generous in my interpretation of Lewis' concept of fiction, perhaps too generous. Lewis, for whatever reason, maintains that *only* in those worlds where the fiction is *told word for word as truth* are worlds do the proper names in that fiction refer – even if all the events depicted occur just as depicted ("Truth in Fiction" 41). This apparently arbitrary limitation on the reference of *functional* proper names can only be made sense of when one considers what sort of



objection Lewis might have been trying to anticipate. There are many stories that are told as known truth in the actual world which did not happen precisely as depicted. In fact, the truism that “every person has his own story” suggests that almost no story ever told as truth depicts the events exactly as they happen. So if Lewis’ standard for a story being fictional is the faithful depiction of the events in a world, then his definition of fiction will end up being fatally infelicitous: nearly every story ever told is fictional by virtue of e.g. exaggeration, figurative speech, metaphor, etc. And if avoiding this pitfall is the only reason that Lewis suggests this constraint (he provides no other reason, anyhow), then the constraint is clearly *ad hoc*. Unless the concrete realist can independently motivate the seemingly arbitrary restriction of referent possible worlds to those in which a given fictional story is *told word-for-word as known truth*, we have a dilemma: concrete realism is either an *ad hoc* theory or it is infelicitous to our concept of fiction.

Perhaps just as absurd is Lewis’ claim that fictional tales *could* be told as truth in another world. Though a concrete realist might conceivably find an independent motivation for the apparently arbitrary restriction of referent possible worlds, it will suffice for our purposes to merely observe that this restriction is utterly implausible. I quote from a random page in *Crime and Punishment*:

Having said this, Svidrigailov suddenly laughed again. It was clear to Raskolnikov that this was a man who was firmly set on something, and who kept his own counsel.  
“You must not have talked with anyone for several days?” he asked.  
“Almost right. And so? You’re no doubt surprised that I’m such a congenial man?”  
“No, I’m surprised that you’re a much too congenial man.”<sup>12</sup>

One need not have any knowledge of *Crime and Punishment* to recognize this passage as fiction. It is not the content that marks it as distinctly fiction to the reader, but rather the *style*. Most

---

<sup>12</sup> Fyodor Dostoevsky, *Crime and Punishment*, tr. Richard Pevear and Larissa Volokhonsky, (New York: Random House, 1992), 284.

stories which are told as truth – at least ones that are as complex as *Crime and Punishment* – are biographies or histories which are riddled with references, guesses, conscious interpretations, and qualifications. In other words, long and complex stories which are told as truth *in this world* are almost always self-consciously constrained to the facts at hand. This sort of constraint is rarely found in fiction works – perhaps only in post-modernist novels such as Mark Danielewski’s *House of Leaves*. Furthermore, there are myriad instances in which Dostoevsky describes the thoughts of a character as if from an omniscient perspective, but such a fiction may only have been told as truth – in any world – by a telepath or a god. Lewis surely was not suggesting that we lean on psychic powers for a concept of fiction. We conclude, then, that this theory collapses when put to use.

## 2.2 Abstract Realism

Though we refer to Peter van Inwagen as an abstract realist about ficta, reduction of ficta to abstracta is only half of the story.<sup>13</sup> Like concrete realism, abstract realism is an adaptation of modal realism: fictional worlds are, in many senses, the same as possible worlds.<sup>14</sup> The first sense in which fictional worlds differ from possible worlds for the abstract realist is that they need not be completely possible – impossible worlds exist also. The second sense in which fictional worlds differ from possible worlds, for van Inwagen, is that the story of modal realism only addresses the disengaged side of the Bifurcation. The engaged side of the Bifurcation is, strangely enough, deflated by means of an anti-realist explanation. But let us explore the abstract realist’s ontology before we consider her act of deflation.

---

<sup>13</sup> Peter van Inwagen, “Creatures of Fiction,” *American Philosophical Quarterly*, 14, 4 (1977): 299-308. Hereafter cited in text as “Creatures”. See also W. R. Carter, “Do Creatures of Fiction Exist?,” *Philosophical Studies*, 38 (1980): 205-215.

<sup>14</sup> See Alvin Plantinga, *Essays in the Metaphysics of Modality*, ed. Matthew Davidson, (Oxford: Oxford University Press, 2003).

For an abstract realist such as van Inwagen, ficta “belong to a broader category of things [called] *theoretical entities of literary criticism*” (“Creatures” 302). Though van Inwagen does not say much about fictional worlds, a glance at his treatment of ficta suggests that one may interpolate a realist Waltonian stance concerning them: fictional worlds are webs of propositions existing abstractly (i.e. not in the physical universe, and not exclusively in the mind, but in their own realm), and the propositions that comprise these fictional worlds are all true. Whether or not van Inwagen actually intended to augment abstract modal realism (and I think he did), this position seems to me the strongest position for an abstract realist, so we will adopt it for the abstract realist. The augmentations to modal realism are threefold: (a) the webs of propositions have a structure, and this structure is one of the subjects of literary criticism; (b) the propositions in the web do not completely determine the fictional world and are thus incomplete, because they are, as a rule, tied to the sentences in the text itself; and (c) there is some measure of impossibility permitted, but we will let literary criticism handle this task as well.

A simplistic construal of abstract realism would maintain that Sherlock Holmes is nothing more than the collection of propositions describing him which can be derived from the stories written about him – a subset of his world. However, this is not the position that van Inwagen takes. For van Inwagen, it would be more accurate to say that the characters in the fictional worlds are the bundles of properties *ascribed to* (van Inwagen’s term) them in the true propositions which constitute the world. For van Inwagen, the primary method of reference to ficta is via *definite description*, and proper names are a secondary method of reference which is parasitic upon the primary (“Creatures” 307). Here we see a distinction between persons in the real world and characters in fictional worlds. A person *possesses* or *exemplifies* a property; whereas, a character *is ascribed* or *encodes* a property. And a property can be encoded only by a

fictional entity (a fictum): an entity  $x$  encodes a property  $F$  iff  $F$  is one among a set of properties which exhaustively constitute  $x$ . A character cannot be a detective in the same way that a person can, because a character just *is* its properties; whereas, a person *possesses* her properties. To say that Sherlock Holmes is a detective is to say that the property being-a-detective is one of the properties which *constitutes* Sherlock Holmes. Again, these properties are exhaustively constitutive of Sherlock Holmes.

Because the abstract realist, unlike the concrete realist, admits the fact of Bifurcation in Speech, our existential class of sentences has a bifurcated interpretation. Consequently, we must split ‘Rodion Raskolnikov’ and ‘Sherlock Holmes’ into two names, each with a different function. Rodion Raskolnikov<sub>1</sub> will be used for the proper name embedded in the text, and about which fictional sentences are spoken. We have already hinted that the abstract realist adopts a deflationary position concerning these usages, so, pending this deflationary explanation, the abstract realist may interpret existential sentences about these engaged usages literally. Rodion Raskolnikov<sub>2</sub> will be used for the proper name when it appears in a disengaged context – that is, in critical sentences. In this context, existential sentences must be interpreted to state that Rodion Raskolnikov<sub>2</sub> is an abstract entity. Hence:

- (F<sub>3</sub>) (a<sub>1</sub>’’) Rodion Raskolnikov<sub>1</sub> does not exist.
- (a<sub>2</sub>’’) Rodion Raskolnikov<sub>2</sub> does not exist except as an abstract entity.

A critical sentence can only be true if it expresses a proposition which is true in the fictional world, and the web of propositions which constitutes the fictional world is derived from (implied by, extracted from, suggested by) the text itself. This web of propositions constituting the fictional world is incomplete, for we are restricted to only the propositions which can be gleaned from the text; nevertheless, its target is maximality. So the set of propositions which constitutes a fictional world is that set which describes the fictional world down to the greatest

detail it possible. For every sentence written by an author, there exists an abstract proposition (or a set of abstract propositions) which partially constitute(s) the world. Using our concept of encoding, the abstract realist paraphrases critical claims thus:

- (F<sub>2</sub>) (a'') Sherlock Holmes<sub>2</sub> encodes being-a-clever-detective to  $x$  degree of cleverness; whereas Porfiry Petrovich<sub>2</sub> encodes being-a-clever-detective to  $y$  degree of cleverness, and  $x > y$ .
- (b'') Sherlock Holmes<sub>2</sub> encodes being-a-famous-detective to  $x$  degree of famousness, and  $x$  is greater than the degree of famousness to which any real person exemplifies being-a-famous-detective.<sup>15</sup>

The only apparent interpretational benefit to being an abstract realist is that analytic critical sentences may be taken *literally*, provided that we define a character as a bundle of human properties. Our previous example, 'Rodion Raskolnikov is an archetypal character' is one such sentence. This sentence may be interpreted literally because a *character* is the sort of entity which we would expect to be composed of properties, for it lacks the implied personhood that a name like Sherlock Holmes bears. A character can only exist in a fictional world; whereas, a detective can only exist in the real world.

The last class of sentences to be interpreted, the sentences which are written by authors and published as works of fiction, "[do] not represent an *attempt* at reference or description" ("Creatures" 301), thus "they are not used by their authors as the vehicles of assertions" ("Creatures" 307). This is because these sentences literally *generate* the ficta about which we speak critically and existentially. If abstract realists were to take these fictional sentences literally, then there would be contradiction between fictional and existential sentences because the only sort of thing that can be a murderer is a concrete entity, namely a person. Hence, we say that for abstract realism, fictional sentences generate characters and fictional worlds, but do

---

<sup>15</sup> We assume, here, that more or less cleverness or famousness can be represented in numerical quantities. While this is certainly a simplification, it still accommodates the intuition that one can be more clever or more famous than another.

not assert anything about them: assertion is saved for critical sentences and existential sentences. Because these engaged sentences do not even attempt to assert, we may take the van Inwagen as a non-cognitivist about the engaged context: engaged fictional discourse is not truth-apt. Thus we have:

### Interpretation under Abstract Realism

- (F<sub>1</sub>) (a'') Rodion Raskolnikov<sub>1</sub> is a murderer. [no truth value, no interpretation]
- (F<sub>1</sub>) (b'') Sherlock Holmes<sub>1</sub> plays the violin. [no truth value, no interpretation]
- (F<sub>2</sub>) (a'') Sherlock Holmes<sub>2</sub> encodes being-a-clever-detective to  $x$  degree of cleverness; whereas Porfiry Petrovich<sub>2</sub> encodes being-a-clever-detective to  $y$  degree of cleverness, and  $x > y$ .
- (F<sub>2</sub>) (b'') Sherlock Holmes<sub>2</sub> encodes being-a-famous-detective to  $x$  degree of famousness, and  $x$  is greater than the degree of famousness to which any real person exemplifies being-a-famous-detective
- (F<sub>2</sub>) (c'') Rodion Raskolnikov<sub>2</sub> is an archetypal bundle of human properties.
- (F<sub>3</sub>) (a<sub>1</sub>'') Rodion Raskolnikov<sub>1</sub> does not exist.
- (F<sub>3</sub>) (a<sub>2</sub>'') Rodion Raskolnikov<sub>2</sub> does not exist except as an abstract entity.
- (F<sub>3</sub>) (b<sub>1</sub>'') Sherlock Holmes<sub>1</sub> does not exist.
- (F<sub>3</sub>) (b<sub>2</sub>'') Sherlock Holmes<sub>2</sub> does not exist except as an abstract entity.

This ontology of abstract realism affords a very satisfactory explanation of the Bifurcation. The fictional sentences that are uttered and written by authors are purely inventive and completely non-cognitive, which reflects the fact that we can never tell an author that he is “wrong” when he writes a sentence: Conan Doyle can have Sherlock Holmes play any instrument he wants him to play. It is this fact that leads van Inwagen to suggest that the action involved in fictional sentences is an *encoding* or an *ascription*. When Conan Doyle writes that Sherlock Holmes plays the violin, he is not asserting it to be the case, for it is not the case until Conan Doyle writes the sentence; rather, he is fashioning a character from nothing. Now that we have posited that fictional sentences do not represent any attempt to assert and therefore have no truth-value, we can accommodate the intuition that a fictional sentence treats fictional worlds as if they are complete: we seem to speak this way only because a fictional sentence does not have

to answer to a truth-value, so there is never a conscious restriction on what may be said about a fictional world in fictional sentences. Conversely, *critical* sentences *do* answer to a truth-value. These sentences are uttered or written *after* the fiction has already been generated, so these genuine assertions are strictly limited to what has already been generated by the (non-assertive) fictional sentences.

But must van Inwagen resist the notion that proper fictional names refer? Must the abstract realist maintain a deflationary position in the engaged context? Nathan Salmon compares this move to buying a Lamborghini only to keep it locked in the garage. If we are going to help ourselves to abstracta, we might as well use them wherever we can – perhaps we can invent a new category of things to which proper names refer in fictional sentences. Suppose, then, that we do posit a set of entities about which fictional sentences assert truths. There are two possibilities: either these ficta are discovered or they are invented. Amie Thomasson suggests that if we are not willing to adopt the position that ficta are invented, then we have some explaining to do about what exactly it is that humans create when they create.<sup>16</sup>

In fact, one of the greatest benefits of the abstract realist's position is that it accommodates the intuition that fictional characters are created. For this reason, let us grant that they are invented. Given this, then proper fictional names either refer or they do not refer. But to what could they possibly refer which exists *before* the character is invented? Two problems fall out of the assumption that names used in fictional sentences refer: (a) The absence of constraints on fictional discourse (as opposed to critical discourse) will complete fictional worlds, thereby threatening to destroy the theory's account of Bifurcation. Fictional worlds cannot be both complete and incomplete, and if we allow names spoken in *both* contexts to refer to the *same*

---

<sup>16</sup> Amie Thomasson, *Fiction and Metaphysics*, (New York: Cambridge, 1998).

entity, we will be forced to take a stance concerning whether they are complete or incomplete. Taking this stance will leave abstract realism with the same dilemma that concrete realism suffered: accommodation of only one side of the Bifurcation. (b) If we grant that proper fictional names refer before we have given fictional characters any properties, then we have essentially granted fictional characters *souls* or *substances* – what could a name refer to before any properties are encoded? Only Locke’s “something I know not what” – yet this directly contradicts the abstract realist’s reduction the character to a bundle of properties. Consequently, we conclude that proper fictional names in fictional sentences do not refer (as one would expect if the sentence also does not assert). Salmon puts the point succinctly:

[O]ur language licenses a certain kind of metaphysical move. It postulates an abstract artifact, the fictional character, as a product of this pretense. But the name 'Sherlock Holmes' does not thereby refer to the character thereby postulated, nor for that matter to anything else, and the sentences involving the name 'Sherlock Holmes' that were written in creating the fiction express no propositions, about the fictional character or anything else. They are all part of the pre-tense, like the actors' lines in the performance of a play. It is only at a later stage when discussing the fictional character from a standpoint outside of the fiction, speaking about the pretense and not within it, that the language makes a second move, this one semantical rather than metaphysical, giving the name a new, non-pretend use as a name for the fictional character (“Nonexistence” 294).

An abstract realist may accommodate this deflationary explication of the engaged context in one of two ways. Either the abstract realist may use a nominalist strategy to reduce apparently referring names to some other speech act, or she may lean on pretense (as Salmon does) as a means of explaining how these sentences do not assert and these names do not refer.

The first alternative has been recently explored in literature about Gappy Proposition Theory,<sup>17</sup> but we do not intend to explore this option here for the second alternative seems a

---

<sup>17</sup> One currently popular method of effecting this nominalist reduction is to use Gappy Proposition Theory. David Braun attempts thereby to show that meaningful propositions can incorporate non-referring or bearer-less names in “Empty Names, Fictional Names, Mythical Names,” *Noûs*, 34 (2005): 596-631. Hereafter cited in text as “Empty Names...” For further discussion of gappy propositions, see also David Braun, “Empty Names,” *Nous* 27 (1993): 449-69. Nathan Salmon, “Nonexistence,” *Noûs*, 32, 3 (1998): 277-319, Hereafter cited in text as “Nonexistence”; Kenneth Taylor, “Emptiness without Compromise,” in *Empty Names, Fiction, and the Puzzles of Nonexistence*, ed. A. Everett, T. Hofweber (Stanford: CSLI Press, 2000), 17-36; and Fred Adams, Gary Fuller, Robert Stecker, “The



more promising avenue for the abstract realist. What the abstract realist will ultimately need out of a concept of pretense is a mechanism which can generate structured webs of propositions from a text full of sentences which are not assertions. The simplest way to accommodate this need is to formulate the semantic mechanics of pretense in a manner parallel to the semantic mechanics of assertions: pretend assertions. Though such a concept of pretense seems intuitively plausible, we leave it to the abstract realist to refine our suggestion and defend it against objection. Our business is to construct a mentalism, so in the name of this task we will simply outline the failures and successes of the two forms of augmented modal realism discussed in an effort to guide and shape our approach toward a satisfying fictional mentalism.

### **2.3 The Failures and Successes of Augmented Modal Realism**

We have seen that concrete realism fails as theory of fiction for three reasons: (a) It cannot explain the intuition that ficta and fictional worlds are created by the human beings who imagine them; (b) it cannot distinguish between fiction and non-fiction without imposing *ad hoc* constraints; (c) it cannot accommodate both sides of the phenomenon of Bifurcation.

It is because concrete realism falters at (a) that it also falters at (b). Lewis' theory cannot take the concept of pretense as anything more than a mental byproduct of fictional discourse, because all of the relevant mechanisms are purported to be explained by possible worlds. More importantly, because nothing is created in fictional discourse, there is no such thing as a pretense.

---

Semantics of Fictional Names," *Pacific Philosophical Quarterly*, 78 (1997): 128-148, Hereafter cited in text as "Semantics".

Consider our fictional sentences (1a, 1b). If 'Sherlock Holmes' and 'Rodion Raskolnikov' do not refer to any entity at all, then we must submit that these sentences have a propositional structure with certain positions left empty: (1a) '\_\_\_ is a murderer,' and (1b) '\_\_\_ plays the violin.' Where we previously had 'Sherlock Holmes' and 'Porfiry Petrovich', we now have gaps in the proposition. The effect that the gaps in these propositions have is to leave the propositional structure unfilled. We can represent the semantic content of (1a) thus:

<\_\_\_, being-a-murderer>

As Braun observes, if a name has no semantic content, then the proposition in which it is used will have incomplete semantic content; thus, gappy propositions are apparently truth-valueless. Hence, under Gappy Proposition Theory, as construed by Braun, fictional sentences are non-cognitive.

Rather, that which we naively call ‘pretense’ is precisely what Lewis is trying to explain in terms of possible worlds. It is for this reason that Lewis suggests that when dealing with *impossible* worlds, we should expect to have a non-trivial concept of truth “only under the pretence – *not to be taken too seriously* – that there are impossible possible worlds as well as the possible possible worlds” (“Truth in Fiction” 45-6, my emphasis). If Lewis takes the notion of pretense seriously, then he runs the risk of undermining his whole project: for then possible worlds *could* be invented rather than discovered. Without a notion of pretense, Lewis can only connect a written story to a possible world in one of two ways: either the story is told word-for-word as truth in a possible world, or it precisely describes the events that occur in a possible world. But if these are his only options, then, as we have observed, Lewis falls into the dilemma of having either an *ad hoc* (and absurd) theory or an infelicitous definition of fiction.

Abstract realism is capable of avoiding all three of the pitfalls through which concrete realism cannot but fail to navigate. So long as abstract realism adopts the intuition that ficta and fictional worlds are *created* and not *discovered*, as Thomasson suggests that it should, it can safely navigate pitfalls (b) and (c). If fictional worlds are, indeed, invented, then the abstract realist is no longer bound rigidly to description or to word-for-word repetition – such an abstract realist does not need to find a way to match what is already in a (discovered) fictional world with what is written in the text. The abstract realist may hold, instead, that the propositions which constitute the world of *Crime and Punishment* are created by the very act of writing the sentences which constitute the text. The connection between the fictional world and the sentences in the text, then, is a purely imaginative connection. Dostoevsky pretended to be speaking the truth about the world, and this act of imagination generated a set of propositions which constitute a fictional world. And here we have both sides of the Bifurcation

accommodated, so in specifically conforming to the intuition in worry (a), the abstract realist is able to also avoid worries (b) and (c).

We conclude that augmentation of modal realism to fit the purposes of a theory of fiction is only capable of addressing one side of the Bifurcation. Concrete realism, a pure realism, must come down one way or another as to whether Conan Doyle imagined a *single* determinate world or a *set* of determinate worlds. Either way, the theory predicts no Bifurcation in Speech – clearly a false prediction. Abstract realism, if it were a pure realism, would reach a similar impediment. If engaged speech acts assert and if ‘Sherlock Holmes<sub>1</sub>’ refers abstractly, then either engaged fictional discourse is identical to disengaged fictional discourse except for an unexplained constraint on what may be said in the disengaged discourse, or Holmes<sub>1</sub> is some other kind of abstract entity which is unrelated to Holmes<sub>2</sub>. In the first case the Bifurcation disappears just as in concrete realism, and in the second case the Bifurcation is so deep that there is no longer any connection between the engaged discourse and the disengaged.

In this second situation, should it be the case that Holmes<sub>1</sub> refers to some new kind of abstract entity, call it a character<sub>1</sub>, and Holmes<sub>2</sub> refers to another kind of abstract entity, a character<sub>2</sub>, then when I say ‘Dostoevsky’s character, Raskolnikov, is a murderer,’ how can I possibly determine whether I am talking about the character<sub>1</sub> or the character<sub>2</sub>? Perhaps it is impossible to speak critically about characters<sub>1</sub>, but why should this be the case since we can speak descriptively of them? Clearly, these troubles would be more easily managed by deflating the engaged discourse in order that the abstract realism may use this engaged discourse to explain how the disengaged discourse becomes *inflated*.

Our discussion of augmented modal realism has posed us with three potential problems besetting a theory of fiction and one potential solution to these problems. If concrete realism

demonstrates the pitfalls that a theory of fiction must be conscious of avoiding, then abstract realism demonstrates that a deflationary concept of pretense can avoid these pitfalls smoothly. And it is important that this concept be a concept of *pretense*, for this is the concept which allows us to avoid all three pitfalls in one swift maneuver. It is important to keep in mind that we have not yet deployed a concept of pretense: thus far, we have only a very basic and broad sketch of pretense, if that. The major stumbling point for abstract realism is the tendency to focus on the realist discourse (the disengaged discourse), and to neglect the anti-realist discourse (the engaged discourse). So before abstract realism may be adopted as a viable candidate, pretense theory must be explored to determine its capacity to support a deflationary discourse. While van Inwagen believes that this discourse is non-cognitive, we need not assume that pretense must adhere to this sort of strong anti-realism. And, in fact, we will discover that pretense is actually a relativistic discourse and is, thus, a weaker form of anti-realism.

Finally, I must stress that nowhere in this essay do I claim to demonstrate that there exists within abstract realism any problem so great that the theory ought to be abandoned in favor of another theory. Though I do support a mentalist anti-realism about ficta and fictional worlds rather than an abstract realism, the task of undermining said abstract realism about ficta and fictional worlds is a task whose ambition exceeds the boundaries of this essay. While I believe that such a feat is possible, I will simply confine my comments to a sketch of this project in my concluding chapter. Because the major theme of this essay is a philosophically constructive essay – an effort to produce and polish a theory of ficta and fictional worlds – we will allow these comments concerning abstract realism to suffice for our purposes.

## CHAPTER 3. PRETENSE THEORY

### 3.1 Waltonian Pretense Theory

Our exploration of augmented modal realism has brought us to the conclusion that a pure modal realism – concrete realism – falls prey to three major criticisms, the foremost of these is the complete shunning of one side of the Bifurcation. A hybrid between modal realism and some sort of pretense theory seems *prima facie* plausible, based on our account of abstract realism. Therefore we will lay out a pretense theory beginning, again, with Walton’s account.

Though Walton does not seem to be aware that the abstract realist posits a realist/anti-realist hybrid in order to explain fictional discourse, he does explicitly develop a deflationary concept of pretense:

It is my contention ... that when realists claim with a straight face that people refer to and talk about fictional entities and that our theory must postulate them in order to make sense of what people say, they are overlooking or underemphasizing the element of make-believe that lies at the heart of the institution. They mistake the *pretense* of referring to fictions, combined with a serious interest in this pretense, for genuine ontological commitment (*Mimesis* 390).

Throughout *Mimesis as Make-Believe*, Walton stresses the fact that when we concern ourselves with fiction, we are participating in an elaborate pretense. We imagine the world to be a way that, in reality, it is not. Just as we imagined ourselves to be cops and robbers in childhood, so now we imagine ourselves to be in nineteenth century Russia observing the psychological fits that a certain dropout student acquaintance of ours is experiencing after committing his first murder.

While Walton’s concept of fiction certainly revolves around the imagination, it also incorporates that which is not imagined: props. For Walton, “props generate fictional truths independently of what anyone does or does not imagine,” and this is because the rules of the fictional world – the “principles of generation” as Walton calls them – dictate that a certain kind of prop is to be treated thus-and-so whether or not the prop *actually* prompts anyone to imagine a

pretense (*Mimesis* 38). Walton's simplest example of a game of make-believe is a pretense in which two children (call them Timmy and Tommy) agree that all tree stumps count as bears. It turns out that there is a "bear" lurking in the forest just a ten feet behind them, but they do not know this and so they do not imagine it, because they have not seen the tree stump which lies ten feet behind them. So while a prop can generate a fictionally true proposition, it cannot generate the principles by which that proposition is made fictionally true – these principles are supplied by those doing the imagining. It is also worth commenting, at least in passing, that an event or an action can do the same work as a prop. So if Timmy pokes a tree stump with a stick, it is fictionally true that a bear has been wounded, assuming that a stick counts as a spear. Essentially, Walton's theory of pretense tells us that real world actions and objects generate truths within a pretense by virtue of the relevant principles accepted by those persons participating in the pretense. According to Walton there are only two tools involved in collaborative imaginings: props and principles (*Mimesis* 39-43). And these principles – agreed upon by all who participate in the game of make-believe – are what *generate* the fictional world from the bare materials of props and events.

The distinction between props and principles already supports the Bifurcation in a robust way, for any feature – any fictional truth – that we imagine within a fictional world which is not implied by at least one principle and at least one prop is necessarily private to us (unless we speak it). So if Tommy starts using rocks to kill bears while Timmy has no idea what rocks are supposed to be in the pretense, Tommy is either cheating in order to kill bears or he is participating in a pretense whose principles are private to Tommy alone. If Timmy and Tommy were pretending to fight against each other, cheating threatens to become a major problem. Should Tommy, who likes to invent principles of generation, invoke the infamous "force-field"

principle which essentially guarantees invincibility, Timmy is likely to respond that force-fields are not allowed in this game. If Tommy refuses to relinquish the force-field principle, the game will be over because a set of principles could not be agreed upon. The disengaged side of the Bifurcation is accommodated by the fact that not just any principle is acceptable in the game of make-believe. The engaged side, on the other hand, is accommodated by the fact that Tommy may still adopt any set of principles he likes, but this lack of constraint entails that Tommy's game will not be a collaborative game of make-believe.

Walton draws the analogy that “what is true is to be believed; what is fictional is to be imagined” (*Mimesis* 41, and that which is “fictional” in Walton's vocabulary is identical to that which is “fictionally true”). Thus, if we are aware of the principles by which a prop is intended to generate a fictional truth, then the prop constitutes a *prescription* to imagine the fictional truth, where “a fictional truth consists in there being a prescription or mandate in some context to imagine something” (*Mimesis* 39). If you and I both read *Crime and Punishment*, there are non-trivial ways in which we imagine the same things. Though Dostoevsky never explicitly mentions it, we generally assume that one day when Raskolnikov is released from Siberia, he and Sofia will live happily ever after (or something like it). The convergence in our imaginative constructs concerning an issue which is only hazily implicit in the text itself will be well explained by reading the normativity implicit in accepted principles of generation into Dostoevsky's description of the world he imagined.

### **3.2 Fictionalist Pretense Theory**

Stuart Brock defend a “fictionalism about fictional characters” which borrows the framework of Waltonian pretense theory with the intention of using it to explain fictional discourse. Like a fictionalist about any other discourse, his method is to propose the

characteristic fictionalist reduction: “to borrow the realist’s paraphrase *P*, without embracing it, and to paraphrase [fictional and critical sentences] instead as ‘according to the realist’s hypothesis, *P*’” (“Fictionalism” 9). This method is supposed to secure an interpretation of our three classes of sentences that is every bit as functional as any realist interpretation, but without all the ontological commitment. So the added benefit of this interpretation is that the fictionalist is able to interpret one of the three classes literally: the existential class. To borrow Brock’s interpretations,

Existential statements are to be treated straightforwardly; they contain no implicit prefix. Fictional statements are statements which are elliptical for claims about the content of a literary fiction. Finally, critical statements are statements about the content of the realist’s theory of fictional characters (“Fictionalism” 9).

Anthony Everett, another fictionalist about ficta, makes use of pretense theory to explain critical and fictional sentences in a slightly more sophisticated way.<sup>18</sup> For Everett, fictional sentences are direct imaginative pretenses in which, for example, Dostoevsky imagines the world to be such that Raskolnikov is a murderer, and when we read *Crime and Punishment* we enter into that same pretense. Everett refers to fictional sentences as  $P_1$  pretenses, and critical sentences, or  $P_2$  pretenses, are characterized in terms of  $P_1$  pretenses:

A  $P_2$  pretense involves our engaging in a  $P_1$  pretense in which we pretend that the world is as it is portrayed in the relevant text and, in particular, that it contains various individuals who have the sorts of properties ascribed to them by that text. But, in addition to this, it will also involve us pretending that these individuals have the sorts of properties which fictional realists attribute to fictional characters. We should, in short, pretend that these individuals have a dual nature (“Against Fictional Realism” 640).

So when I assent to analytic critical sentences like ‘Rodion Raskolnikov is an archetypal character,’ I am betraying the  $P_1$  pretense in which we all engage by describing Raskolnikov as an “archetypal character” – for *within the fictional world*, Raskolnikov is not a character at all: he is a person. For the fictional fictionalist, however, another pretense is involved in critical

---

<sup>18</sup> Anthony Everett, “Against Fictional Realism,” *Journal of Philosophy*, 102 (2005): 624-49. Hereafter cited in text as “Against Fictional Realism”.



discourse. Where the abstract realist only invokes pretense theory to explain fictional sentences, the fictional fictionalist invokes pretense theory to explain *both* fictional sentences *and* critical sentences. Because it is the abstract realist's paraphrase which is borrowed as a pretense for the critical discourse, the abstract realist's pretense account of *fictional* discourse is also expropriated in full. The two pretenses – the abstract realist's pretense about persons, and the fictionalist's pretense about characters (posited by the abstract realist outside of the fictional pretense) – are distinct from each other in fictional fictionalism, for the pretense adopted from a disengaged stance, the realist's pretense, pretends *within it* that there also exist persons (who do not actually exist). Clearly, for the fictional fictionalist, the fictional pretense is parasitic upon the critical pretense, for it is within the critical pretense that the fictional occurs. So in addition to the pretense in which Raskolnikov is thought of as a person, I am engaging in a second pretense – the realist's pretense – which pretends that there exist such things as characters, plots and other entities posited by critical fictional discourse.

The primary difference between Everett's treatment and Brock's is that Everett is aware that switching back and forth between pretenses is unavoidable in critical fictional discourse. Everett's distinction between  $P_1$  and  $P_2$  pretenses echoes the conclusion that we reached earlier in response to Sorenson's fictional sorites paradox: the dual nature of fictional characters is another way of describing the Bifurcation in Speech. The existence of fiction sorites paradoxes showed us that we cannot escape the essential difference between speaking about the Sherlock Holmes who is a man in the fictional pretense, and the Sherlock Holmes who is merely a character in the critical pretense. Consequently, a fictional fictionalist's interpretation of the three classes of sentences must distinguish between three usages of 'Holmes'. Let 'Holmes<sub>1</sub>' indicate a reference to Holmes in a  $P_1$  pretense and 'Holmes<sub>2</sub>' indicate a reference to Holmes in a  $P_2$  pretense. In the

first usage, we pretend Holmes is a person and in the second usage we pretend that Holmes is a character. We will use ‘Holmes’ without a subscript to indicate reference to both fictional entities: the person and the character. This last usage is reserved for existential sentences which operate entirely outside of any pretense. Thus:

### Interpretation under Fictional Fictionalism

- (F<sub>1</sub>) (a'') Rodion Raskolnikov<sub>1</sub> is a murderer in  $w_{CP}$ .
- (b'') Sherlock Holmes<sub>1</sub> plays the violin in  $w_{SH}$ .
- (F<sub>2</sub>) (a'') According to the realist’s paraphrase, Sherlock Holmes<sub>2</sub> encodes being-a-clever-detective to  $x$  degree; whereas Porfiry Petrovich<sub>2</sub> encodes being-a-clever-detective to  $y$  degree of cleverness, and  $x > y$ .
- (b'') According to the realist’s paraphrase, Sherlock Holmes<sub>2</sub> encodes being-a-famous-detective to  $x$  degree of famousness, and  $x$  is greater than the degree of famousness to which any real person exemplifies being-a-famous-detective.
- (c'') According to the realist’s paraphrase, Rodion Raskolnikov<sub>2</sub> is an archetypal bundle of human properties.
- (F<sub>3</sub>) (a'') Rodion Raskolnikov does not exist.
- (b'') Sherlock Holmes does not exist.

Steven Yablo has invented a similar semantic apparatus which he uses to extend pretense fictionalism to *all* realist discourses, fictional discourse included.<sup>19</sup> This mechanism underlies what Yablo describes figuralism – a reflexive pretense theory which Yablo intends to be understood as being analogous (if not identical) to figurative speech. He begins with an examination of the truth conditions for any sentences interpreted literally in the English language:

Now, the rules of English make their contribution roughly like so. The rules tell us which sentences are true under which worldly conditions. If  $K$  is a condition sufficient by  $R$ ’s lights for the truth of  $S$ , we write  $R^K \geq S$ . If  $K$  is necessary by  $R$ ’s lights for the truth of  $S$ , we write  $R^K \leq S$ .  $R^K = S$  then means that  $K$  is exactly what is needed for  $S$  to come out true, where truth is judged according to  $R$  (“Go Figure” 76).

If a sentence  $S$  is true within the context of a rule  $R$  given condition  $K$ , then the literal content of  $S$  is  $K$ , for  $K$  is what fixes the truth value of  $S$ . Yablo interprets this mechanism in terms of

---

<sup>19</sup> Steven Yablo, “Go Figure: A Path through Fictionalism,” *Midwestern Studies in Philosophy* 25 (2001): 72-102. Hereafter cited in text as “Go Figure”.

pretense. Thus, the real content of a fictional sentence S is the K which makes S true under a pretense F. Hence we have,

$$(11) \text{ realcontent}(S) = (\text{the } K \text{ such that } F^K = S).$$

Yablo uses logarithm notation to express this formula:

$$(12) \text{ realcontent}(S) = \log_F(S) = K.$$

His examples of the formula in action are as follows:

$$(13) \log_{F\text{-Number}}(\text{the number of } E\text{'s} = n) = \textit{there are } n \textit{ } E\text{'s}.$$

$$(14) \log_{F\text{-Modal}}(\text{there is a world such that } H) = \textit{possibly } H.$$

$$(15) \log_{F\text{-Property}}(x \text{ has } Q\text{-ness}) = x \textit{ is } Q.$$

The key notion is that  $\log_F(S)$  is the fact that makes S *fictional*, and *not* the fact that makes it *true*. If S were true, it would not be a fictional sentence. Suppose that S is: ‘There is a world such that H.’ In assenting to this sentence, we are effectively assenting to the truth of the proposition that ‘S is fictional,’ and the content of the sentence is *possibly H*. What is curious about this formula is that F is a variable, so there is not, in principle, only one fiction under which any given sentence is true. This suggests that the meaning of a sentence is not fixed to any given fictional world. Yablo takes this to mean that the meaning of sentences within a pretense is independent of that pretense (though, seemingly not independent of *any* pretense).

Yablo tinkers with this apparatus numerous times throughout his essay to end up with a concept of pretense which is very similar to Everett’s, so we will simply limit our mention of Yablo’s brand of fictionalism to his conclusion: all cases of figurative speech are, in fact, cases of fictional discourse. Hence, when I say that I have butterflies in my stomach, I am not inviting questions as to what kind of butterflies they are; I am speaking a sentence according to a fiction, and that sentence’s interpretation is constrained by the rules of that fiction:

$$(16) \log_{F\text{-Butterfly}}(\text{there are butterflies in my stomach}) = \text{my nervousness is upsetting my stomach.}$$

In Waltonian terms,  $\log_{F-Butterfly}()$  is a principle of generation. ‘my nervousness is upsetting my stomach’ is the real world event which the principles of generation connects to the fictional truth ‘there are butterflies in my stomach’. Consequently, insofar as a fiction is figurative, the real world content that is expressed in the fiction count as the conditions which make the fiction fictional. *Prima facie*, this seems to be a problematic account of fiction. A figuralist fictionalist assumes that there is some real-world content to every fictional story – that is the figuralist assumes that every story has a “moral”. Such a strong claim might fall prey to fantasies devised for the sake of pure entertainment. Nevertheless, intuition tells us that if a story has no connection to the real world, we would not be able to comprehend the story. Incorporating figuralism into Walton’s account, we find that the moral of the story – the real-world content which renders a figurative sentence fictionally true – apparently *acts as the prop* in a work of fiction. But if this were the case, then the literary critic’s primary concern would be with props. The real content of a figurative sentence is not the prop which generates the fiction, for the tree stump is not part the real content of the make-believe game that Timmy and Tommy are playing. What Yablo is concerned with when he describes the real content of a figurative sentence is *meaning*. Based on this notion, we can extend our concept of pretense a bit further: A pretense is generated from principles which prescribe certain imaginings which are built upon real-world props, and the meaning of this pretense is the real-world content which is figuratively represented.

What, then, does Timmy’s and Tommy’s game of make-believe mean? This question is actually a mistaken question, for Timmy and Tommy play for fun, without any notion of meaning. They are exercising their capacities for imagination, they are enjoying entertainment. This suggests that figurative speech is actually a *subset* of fiction, rather than the other way

around. There is, however, meaning in *Crime and Punishment*, and the benefit of Yablo's pretense mechanism is that it explains how it is that *Crime and Punishment* has meaning.

But there is still one last question that must be answered: in our narrow case of canonical works of fiction, what are the *principles* and what are the *props*? What real world objects and events generate truths within a fiction as outlined by what rules? Walton's view on the matter is that a typical fictional text (like *Crime and Punishment*) is a representation: it *describes* a fictional world *imagined* by the author and *prescribed* to the reader (*Mimesis* 353-5). The description – the text – might be taken as a long string of principles about how we are to imagine (viz. in what order we are to imagine, in what language we are to imagine, borrowing which actual objects as further props, etc.), for it is the text that constrains our imagination in the disengaged context, just as it is the principles of generation that constrain our imagination in a game of make-believe.

What, then, is the prop? One might think that the prop is the book itself, but the book itself is a very poor device for serving as a prop. It would not serve very well to imagine the book to be a bear, or a spear, or Raskolnikov himself. Perhaps it will help to consider an artwork similar to a work of fiction: theatre. In an instance of Shakespeare's *Othello*, actors, actresses, clothing, sets, fake swords, and all the words and deeds of the actors and actresses function as props. The principles of generation, on the other hand, were written as parentheticals in the original text, and they were meant not to be spoken, but followed by the director. Consequently, the principles by which our make-believe game of *Othello* is generated are united with the props which generate truths according to those principles within our game.

Based on an analogy with theatre, we might think that that the text is, after all, *both* a prop *and* a set of principles. But let us resist this thought. If, in a work of fiction, what is to be

imagined is an *entire world*, then the most natural prop to imagine *as* this fictional world is the actual world. Consequently, the events and objects which constitute the actual world become a prescription to imagine a fictional world in which actual objects and events (such as Petersburg, London, Napoleon's reign, etc.) are linked to the objects and events within our pretense via the text of the fiction we are reading.

The need to secure a set of props in the actual world becomes even more urgent when we consider that the experience of an imaginative construct is private. As such, if we are to have a meaningful conversation about a fictional world, or if we are to have some means of collaborating in an act of make-believe, then we will need some way to know that we are talking about the same things. But for those of us who have never been to Petersburg, Russia, a sufficiently robust set of props which fix the propositions that are true in the world of *Crime and Punishment* will be hard to find. In fact, it is not the *actual* Petersburg which serves as a prop in *Crime and Punishment*, for if it were the book could only be read in Petersburg! Rather, it is the reader's *impression* of Petersburg which serves as a prop. Consequently, our earlier conclusion missed the mark: it is not the actual world which serves as a prop for a fictional world, but our impression of the actual world. Dostoevsky's work gives us a set of principles by which we are prescribed to configure our impressions of the real world in order to produce the fictional that is *Crime and Punishment*. So although Walton's phenomenology has given us a useful description of the act of make-believe, it does not answer our questions about the ontology of pretense; it merely converts the question into one concerning imagination and impression. We are apparently lead to a mentalism about pretense – still an anti-realism, but not the kind of anti-realism that the pretense theorists expected. However, we must stave off this path of inquiry for the moment, in favor of raising doubts about pretense theory – for there are doubts to be had.

### 3.3 Objections to Pretense Theory

#### 3.3.1 No Systematic Set of Principles of Generation

There are five objections to pretense theory which I want to consider. The first four are objections raised by Jason Stanley,<sup>20</sup> and the last is my own. Each of these objections will be answered within the purview of this essay – some in the present section and others in later sections. Ultimately, I take these objections to demonstrate that pretense theory, as expounded above, is insufficient to the task of a theory of fiction, though its failure suggests that a mentalism which borrows many of the features of pretense theory may succeed where pretense fails.

The first objection is an objection to the apparent ubiquity of figurative speech according to the figuralist. If fictional discourse is to be interpreted figuratively, then all expressions are, effectively, idioms: each figurative expression is to be understood in terms of its own principles of generation, and these are not reducible to a systematic set of principles. Stanley observes that even Walton apparently agrees on this point: “‘It is fictional that I speak truly in saying ‘N doesn’t exist’ if N-ish attempts to refer fail, for whatever reason.’”<sup>21</sup> If negative existential sentences fail to refer, then even these apparently literal sentences are figurative, which suggests that any discourse which has the trappings of literality may be at bottom figurative.

Figurative interpretation of apparently literal discourses is problematic because compositional semantic theories (in which the semantic value of the parts of a sentence fix the semantic value of the entire sentence) have been seen as solving a major intuitive problem. We somehow have the ability to understand unique sentences with which we have no previous

---

<sup>20</sup> “Hermeneutic Fictionalism,” *Midwest Studies in Philosophy*, 25 (2001): 36-71. Hereafter cited in text as “Hermeneutic Fictionalism”.

<sup>21</sup> “Existence as Metaphor?” In *Empty Names, Fiction and the Puzzles of Non-Existence*, ed. Everett and Hofweber (Stanford: CSLI Press, 2000), 84. Quoted by Stanley in “Hermeneutic Fictionalism,” 40.

familiarity; furthermore, in discourses such as mathematics, we have no reason to think that the sentences used are *not* compositional. In addition to these worries, if there were no systematic set of principles of generation, our ability to understand new and unique sentences would suggest that we already know an infinite set of principles of generation. As Stanley has it, "The importance of compositional semantic theories is that they answer a mystery about linguistic understanding: how can a finite mind grasp the real world truth-conditions of an indefinite number of new sentences?" ("Hermeneutic Fictionalism" 44).

Despite the figuralist fictionalist's commitment to a figurative treatment of the existential class of sentences, this fact does not entail that *all* apparently literal discourses are to be treated figuratively. In discourses such as mathematics, what is interpreted figuratively is the *basic* pretense, the  $P_2$  pretense. According to this pretense, what is pretended – what is spoken figuratively – is that there are such things as numbers. But this figurative treatment of mathematics is nothing more than the existential class of sentences which are relevant to the mathematical fiction. Consequently, we have reason to believe that the *only* kind of apparently literal sentences which are to be interpreted figuratively is the existential class of sentences. Sentences which occur in the mathematic  $P_1$  pretense are to be interpreted compositionally, within the pretense that there are such things as numbers. Similarly, a compositional treatment of traditional fictional discourse may be had (if needed), though only within the basic pretense that there are such things as ficta. Thus, there *can be* a systematic set of principles of generation in  $P_1$  pretenses.

### **3.3.2 Pretense Theory Is Apparently Not Hermeneutic**

The second objection assumes the success of the first objection. According to this objection, if all fictional discourse is figurative – that is, if there is no composition semantic



theory which can fit into figuralist fictionalism – and there is no discourse which is interpreted literally, then apparently figuralist fictionalism is revolutionary and not hermeneutic.

Hermeneutic fictionalism about discourse *D* claims that all sentences spoken in *D* already function as the fictionalist claims they do. That is, the hermeneutic fictionalist thinks of herself as describing what we *already do*. The revolutionary fictionalist, on the other hand, admits that sentences in *D* entail ontological commitments, but that we should cease to speak this way and adopt the fictionalist’s ontologically lighter structure. Consequently, given that the figuralist fictionalist must interpret apparently literal sentences figuratively, it appears that – despite his claim otherwise – the hermeneutic fictionalist is actually a revolutionary fictionalist. This objection is relevant to our purposes for two reasons: (a) the first and most fundamental function of pretense theory is to describe what we *already do* when we speak fictional sentences: a hermeneutic project; (b) the hermeneutic is the more compelling position of the two.

While this objection is contingent upon the first objection, which we believe can be adequately met, the question whether fictional fictionalism is revolutionary or hermeneutic is a relevant one, so let us not forget this particular complaint.

### **3.3.3 No First-Person Authority Concerning Beliefs**

"Consider hermeneutic fictionalism about arithmetic. Competent users of arithmetical discourse will certainly deny that they are pretending when they discuss arithmetic. In such cases, the hermeneutic fictionalist must maintain that the fact that the language user is pretending is not accessible to her, even in principle" ("Hermeneutic Fictionalism" 46). Similarly, we may say that competent users of the English language deny that they are pretending when they speak negative existential sentences about ficta. But the figuralist claims that there is a basic pretense involved in the sentence ‘Sherlock Holmes does not exist,’ namely the pretense that there is such

a thing as Sherlock Holmes about which we can speak. Stanley observes that "this introduces a novel and quite drastic form of failure of first-person authority over one's own mental states" ("Hermeneutic Fictionalism" 47).

This claim, however, seems relatively trivial: the white-power supremacist is not aware of being a bigot; the fundamentalist is not aware of being closed-minded; many Nazis were not aware that Jews were human beings. Stanley admits that such a failure of first-person authority concerning beliefs is akin to Freudian psychological repression, but his objection to such an interpretation is confoundingly weak: "[the hermeneutic fictionalist] probably does not wish to commit herself to the view that the non-introspectible nature of ... pretense is due to childhood trauma. The hermeneutic fictionalist about arithmetic certainly should avoid defending the view that arithmetic is a mass neurosis" ("Hermeneutic Fictionalism" 47). I see no reason why we should assume that repression is universally caused by childhood repression, except out of respect for Freud as an authority concerning matters psychological. Furthermore, the hermeneutic fictionalist ultimately *does* claim that *realism* about arithmetic (and consequently also realism about ficta) *is* a mass neurosis. Stanley has apparently supported his objection with a word-game. He has equivocated between the *discourse* of arithmetic (and fiction) and *realism about the discourse* of arithmetic (and fiction). The hermeneutic fictionalist surely should not hold that the first is a neurosis, but holding that the second is a neurosis is really quite trivial considering that the realist likely holds that the hermeneutic fictionalist position concerning these discourses is also a neurosis.

### **3.3.4 Susceptibility to Empirical Defeat**

Autistic persons do not participate in games of make-believe because they do not understand the purpose of them, but they *do* participate in apparently non-figurative discourses

such as mathematics. Stanley observes that the pretense theorist may maintain that an autistic person can still do mathematics because they take themselves to be speaking literally and not figuratively. But if this is the case, Stanley complains, then it seems that there are *different* psychological processes are involved in make-believe than are involved in mathematics.

Supposing that there are different psychological processes involved in make-believe than are involved in mathematics, our reasons for grandfathering mathematics into the context of make-believe seem to be undercut.

This objection does not seem to have direct relevance to the fictional fictionalist, but there is indirect relevance. If there is an anti-realist concept of fiction available to the fictionalist, then it is this concept of fiction upon which the fictionalist must lean in supporting her broader project of explaining all apparently ontologically committal theories in terms of fictionalism. In short, a good and forward-thinking anti-realist will want her theory of fiction to be sufficiently robust to support all the uses to which she may later put the theory. Consequently, we will want to preserve the possibility of an unproblematic notion of mathematical fictionalism.

If we have been able successfully to meet objection #1, then we may meet this objection equally successfully. In our response to objection #1, we observed that Stanley does not respect the distinction between  $P_1$  and  $P_2$  pretenses, or engaged and disengaged pretenses, respectively. In an engaged pretense, we assume a pretend ontology which is granted to us in the disengaged pretense, and within this ontology we maintain compositional semantics for discourses which are apparently compositional. That is, some discourses are figurative *only* in the disengaged context; whereas, other discourses are figurative in *both* the engaged *and* the disengaged context. Thus, the difficulty that autistic persons have lies in the figurative engaged or  $P_1$  pretenses. Such an autistic person can apparently only engage in  $P_1$  pretenses which are non-figurative. However he

will be able to engage in  $P_2$  pretenses which are figurative. But such an autistic person would apparently have no first-person access to this  $P_2$  pretense, because he does not understand its figurative nature. Thus, the autistic mathematician is stuck in the realist neurosis concerning mathematics.

### 3.3.5 Question-Begging Ontology

It is not enough to simply claim to be able to borrow a realist interpretation and reject realist ontologies of ficta. Intuition suggests to us that there is something amiss in this way of thinking.<sup>22</sup> If we accept Occam's razor as a basic principle, then the reason that the realist adopts his ontology of ficta (whether concrete or abstract) is *entirely* for the sake of the explanation.<sup>23</sup> Fictional fictionalism, on the other hand, tells us that we can have the fictional cake and eat it too. Brock has downplays the fact that the concept of fictionalism *presupposes the concept of fiction* in favor of using it to generate interpretations of the three classes of sentences. For if the realist's paraphrase is itself a fiction, then Brock's paraphrase should really be worded: 'according to the realist's *fiction*, P.' Thus, fictional entities and fictional worlds are themselves fictions, and so much is suggested by Brock himself: "one could hardly be a realist about fictional characters on the one hand, and a[n anti-realist] fictionalist about some alternative metaphysical theory, *T*, on the other" ("Fictionalism" 11). We will have to be very careful in dealing with a notion like this, for at all junctures we run the risk of begging the question against ourselves by assuming a concept of fiction in our definition of the concept of fiction.

Does Yablo's pretense mechanism advance us toward the fictional fictionalist's goal of ontological emptiness? Not yet, but he is headed in the right direction. He has created a paraphrase of sentences that seem to entail the existence of an entity which transforms them into

---

<sup>22</sup> And so does Quine, who "deplore[s] the philosophical double talk, which would repudiate an ontology while simultaneously enjoying its benefits" *Word and Object*, (Cambridge, Mass.: MIT Press, 1960), 242.

<sup>23</sup> Though in practice this is probably not always the only motivation.

sentences that entail only the existence of a fiction. Ficta, as we supposed earlier, piggyback upon fictions. Yablo also comes to the conclusion that a bifurcation in speech is necessitated when using fictions: an anti-realist sentence like “the number of numbers is 0” cannot be reconciled without assuming an equivocation (“Go Figure” 80). Thus, the first instance of ‘number’ occurs *within* the fiction (and is thus spoken in an engaged fashion) and the second instance occurs *without* (and thus spoken in a disengaged fashion). But, as we have stressed numerous times throughout this essay, the Bifurcation is *not* an explanation; it is a *fact* that needs explaining.

When we speak in a disengaged fashion, we play but a single language game, and we interact with our fiction from the safe confines of our (presumably nominalist) meta-language. Yablo calls this a *basic* language game: “G is *basic* if acceptability in G is a function of how things really are” (“Go Figure” 83). But when we engage in the pretense – when we speak about fictions from within – we must play a new language game which is *parasitic* upon the first: “G\* is *parasitic* if acceptability in G\* depends on how things are imagined to be when playing some other game (as it might be, G)” (“Go Figure” 83). In order to play a fictional language game, we must first know both the extent to which it is fictional and the rules which govern the game. When we say that ‘numbers do not exist’, we are playing a basic language game: we pretend that there are such things as numbers, and we state that they do not exist. But when we say that ‘the number of numbers is 0’, we are playing both a basic and a parasitic language game: the first usage of “number” occurs within the parasitic language game – the game of mathematics which explores the relationship of numbers to each other; the second usage of “number” occurs within the basic language game – the Platonist game which pretends that there exist such things as numbers. So in the basic language game, we relate the fiction to the actual world; in a *parasitic*

language game, we participate in the fiction, assuming that in the background there lies another language game which relates the game to the actual world.

The bifurcated within/without engaged/disengaged structure of figuralism or fiction (as you prefer) is a fact that is embedded in the human practice of creating fictions and using figures of speech. Again, this fact existed prior to philosophy, so it is a fact that needs explaining. In “Go Figure,” Yablo progressively works toward this figurative model in his attempt to invent a form of fictionalism which is sufficiently robust to be able to explain away all ontological claims that a fictionalism might want to explain. What Yablo downplays in this inventive process is that he is relying, the entire time, on a concept of fiction: according to Yablo, the fact that the use of modal logic entails the existence of possible worlds is the very fact that makes the sentence “there are possible worlds” *fictional*, rather than *true*. The only context in which a fictional sentence is true is *within* the fiction, and this only occurs if the discourse in which the sentences is used approves of the sentence. In other words, anytime we seem to be incurring the existence of some entity in our speech, we are really creating for ourselves a simulation, a fantasy, a fiction, in which these entities exist – we are not required to believe that these entities exist beyond the fiction.

In order to understand the different levels of pretense, it will be useful to bring in Carnap’s internal/external distinction: the entities entailed by a discourse exist only *internal* to the discourse; whereas, *externally*, the discourse and its entities are but a pretense. So, again, of *what* does a pretense consist? I submit that we have a reason to reject the notion that there are no external claims involved in fictionalist explanations. Let us begin with the discourse of physical science. The facts experienced about the material world are explained beautifully – though not completely – by the scientific model of the material world. We may, if we like, construe the

scientific model as a pretense, but the structure of pretenses requires that any participant in the scientific pretense must at the same time admit the existence of something external to the pretense in order to distinguish the pretense as such, namely: the material world which is modeled by the scientific pretense. In taking the scientific model as a pretense, we believe in the existence of such entities as electrons and quarks only insofar as this belief aids our understanding of the material world. I.e. we pretend that the material world is composed of these entities. But in so pretending, we have already assumed Carnap's distinction: Our commitment to the belief that quarks exist is entirely *internal*, for it is subject to change in the case that it fails as an explanation of what we take to be *external* to the pretense.

Carnap's distinction will, however, drop out if I take scientific discourse as constitutive of the material world: in the absence of pretense, there is no Bifurcation. So it is only insofar as I am *aware* that the language game I am playing is *only a game or a pretense*, that a question of externality has any relevance. If I do not even know that I am participating in a pretense, I will approach my subject with a seriousness reserved for the external. So, for example, in the case that you tell me that you have butterflies in your stomach and I do not realize that you are speaking figuratively, I might ask you why you ate butterflies. In asking this, I am ignoring *both* language games, not merely the external language game. This is what is known as ruining the game or killing the joke. Accordingly, Carnap's internal/external distinction constitutes the necessary and sufficient conditions for an act of pretense and is therefore integral to the notion of pretense. So while our scientific explanation of the material world (if we accept figuralist fictionalism) need not demand any external obligations concerning *material* entities, it does demand external obligations concerning *pretenses* about material entities. If it is one of the defining marks of an imaginative construct that it has no features other than those that are

necessary for the explanatory use to which it is put (e.g. an electron has “no hairs” as the physicists say), then phenomenon of pretense seems to escape its own rubric! Whatever it is that a *fiction* or *pretense* is, this (possibly abstract) entity seems to be more of a *discovery* than an *invention*. This, however, is still only an intuition – the intuition that it is not enough to simply describe a piece of empirical data on the one hand and then to claim that it has no source on the other.

Suppose our pretense theorist desires to evade an external ontological commitment to pretenses. He will be inclined to construe the internal/external distinction in terms of pretenses: the only thing external to a pretense is – you guessed it – yet another pretense. However, every pretense is built upon the internal/external divide, so if our theorist seeks to explain a pretense with a pretense, he has a dilemma on his hands: either there is a regress of pretenses (whether circular or linear), or there is a master-pretense which defies the internal/external bifurcation.

In the first case, we are committed to the existence of a web of pretenses, and we are hopelessly caught in this web. For example, when I say that Sherlock Holmes is a detective, I am pretending that there is such a person. External to this pretense is the pretense that there are characters. External to this pretense is the pretense that there are pretenses. And so on.

In the second case, we truncate the regress by invoking *ad hoc* the *existence* of a master-pretense which does not rest upon an internal/external distinction. This move does not seem to be any different from claiming that there is an external world (the master-pretense), and in it there exist pretenses (subordinate pretenses). In both cases, we have what looks and feels like an external ontology of pretense. This act of placing pretense at the heart of all experience will lead us to a conclusion similar to Descartes’ *cogito*: “I pretend, therefore I am” or, to construe the



claim in a less question-begging fashion, “There is pretending”. And this is an ontological claim if there ever was one.

## CHAPTER 4. COLLINGWOODIAN MENTALISM

### 4.1 Naive Mentalism

Our goal is to construct an anti-realist ontology of fiction which is consistent with Walton's concept of fiction as pretense/make-believe. Based on our consideration of fictionalism, pretense and make-believe theory entail some sort of ontology – namely an ontology of pretense. Similarly, Walton's notions of principles of generation and props, once applied to works of fiction, rely on imaginative impressions of the actual world. Both of these consequences suggest that pretense theory – must be bolstered by a mentalism. I should first make sure it is clear that, minimally, anti-realism (as defined in this essay) entails only that sentences in a discourse are not true (or false) independent of human cognition – it does not entail the nonexistence of truth-making entities. The silence of pretense theorists on this issue implies that an anti-realist theory of fiction somehow assumes the nonexistence of pretenses, but this position is unnecessarily extreme. The mind-dependence of pretenses need not suggest nonexistence, but if we do reject the nonexistence of pretenses, then we must adopt a theory under which a pretense exists as mental construct – a thought. Our mentalism begins with the claim that a fictional world is a pretense and a pretense exists in a human mind as a complex thought.

First, our mentalism will borrow the semantic machinery that fictionalism uses to reduce all ontological questions to questions about pretenses – questions about entire fictional worlds. This entails that we maintain Yablo's and Everett's notion that pretenses or fictional worlds are reflexive and come in pairs (at least): one an ontological pretense in which all the necessary critical entities are posited, and one a fictional pretense in which a world like ours is posited.

The intuition of creativity can be accommodated simply and appealingly, for all mental constructs are created by the person who imagines them.

The intuition that the negative existentials are to be interpreted literally can only be handled by mentalism if “existence” is taken to mean “realist existence” or “mind independent existence”. Because ficta piggyback upon fictional worlds, all the ficta which populate Sherlock Holmes’ fictional world exist – but only in the mind of the reader. This explanation of ‘Sherlock Holmes does not exist’ is still faithful to the spirit of the sentence, because *my* Holmes does not exist for anyone but me. So anyone who speaks the sentence ‘Sherlock Holmes does not exist’ speaks it for the sake of those who hear it, and not for her own sake. Such an interpretation of negative existentials completely eliminates the problem of empty names (which is tied to the first objection raised by Stanley against pretense theory), because all empty names can be reduced to names which refer to mentalistic ficta.

One of intuitions involved in the Bifurcation of Speech suggests to us that within Sherlock Holmes’ fictional world, Holmes has a definite number of hairs on his head. In fact, the entire world is uniquely determined, because Sherlock, or Mycroft, or Watson can empirically verify whatever it is that the texts have left underdetermined or elliptical, for they are the ones who inhabit the world. That is, within the pretense of Holmes’ world, the number of hairs on Holmes’ head can theoretically be counted. While this intuition persists, it is one that a mentalist anti-realist cannot accommodate: if ficta and fictional worlds exist only in human cognition, then they are limited by the finitude of the human mind. The fact that we have this intuition must be explained away by our ontology.

In our literary criticism of *Crime and Punishment*, the incompleteness that we experience is a direct result of our intended faithfulness to the fictional world which Dostoevsky established

(described, represented) within the text itself. It is likely that Dostoevsky imagined Raskolnikov to have a face, and that this face had distinct features in Dostoevsky's mind. This does not entail that the face of Raskolnikov is *incomplete* in the fictional world, because a mere act of imagination can provide Raskolnikov with a face; rather, the face of Raskolnikov is incomplete in Dostoevsky's *depiction* of the fictional world that he had in mind (in which Raskolnikov likely had a face). When I read *Crime and Punishment*, I cannot help but imagine Raskolnikov to be a gangly man with a face whose features are markedly attractive but sharply defined. When I imagine the thick eyebrows, thin lips, average ears and close-cropped black hair, I am filling in the details that Dostoevsky left out for the sake of brevity. I do not intend to do violence to Dostoevsky's vision, so when I engage in critical discussion I do not mention the features that I have projected onto Raskolnikov's face. Instead, I mention the properties that Dostoevsky either explicitly or implicitly ascribed to Raskolnikov *in the text*. The fact that the Raskolnikov that I imagine is different from the Raskolnikov that Dostoevsky imagined or the Raskolnikov that you imagine is not problematic, because there is an enormous catalogue of features that the two Raskolnikovs can share: viz. those features which have been left unmentioned in the text itself.

According to our naïve mentalism, when I say 'Raskolnikov' I refer to something completely different from what you refer to when you say 'Raskolnikov'. While our references differ, the degree to which your pretense is *similar* to my pretense is just the degree to which Dostoevsky left the description of *his* pretense incomplete. This is why I am not upset when you tell me "No, that's not how I pictured him at all!" Thus, the difference between my pretense and your pretense is trivial as far as the actual text is concerned. Therefore, let us tentatively say that ficta are mentalia whose trivial properties are purely subjective and whose important properties have been specified as deemed appropriate by the author. And many of the unspecified features

of *Crime and Punishment* can be easily filled in by analysis of the culture in Petersburg, Russia at the time. But even with complete access to cultural context and the full text of the book, one can still find myriad unspecified features.

But how can I know that the difference between my Raskolnikov and your Raskolnikov is trivial? Generally speaking, among elliptical features of the world established in *Crime and Punishment*, there are two types: the trivial and the non-trivial. The non-trivial elliptical features of an established fictional world are those features that we have hitherto described as indeterminate. For example, in Tatyana Tolstaya's novel *The Slynx*, the Slynx is a mythical creature, but the question is never answered whether the proposition 'the Slynx exists' is true within the fictional world established in the text. So the existence of the Slynx is a non-trivial elliptical feature of the fictional world established by Tolstaya. An elliptical feature is trivial if its hypothetical determination contributes no useful literary criticism of the work. Thus, the number of hairs on Sherlock Holmes' head is a trivial elliptical feature of the fictional world established by Conan Doyle.

While literary criticism is bound to attempt to fill in the non-trivial gaps in the fictional world (an act which most readers engage in on their own), filling in some of the trivial gaps becomes a subjective practice that most of us find necessary when we attempt to imaginatively construct the fictional world of the text. Hence, the bifurcation we mentioned earlier, Everett's "dual nature" of fictional characters, is a consequence of the fact that in order to imagine the fictional world depicted by any given work of fiction, we must imaginatively determine certain indeterminate features of the fictional world – for in order to pretend that the fictional world is the real world, this fictional world must be taken to be – at least in principle – just as determinate as the real world.

But *must* we imagine Raskolnikov to have a face? Probably not, for we certainly do not imaginatively determine within our pretense every single feature which is in-principle determinable. So then why do we speak as if fictional worlds are complete? The answer to this question is much simpler than one would think: as mentioned in Section 3.1, the act of make-believe is a *game*, and while everyone who plays the game is aware that it is a game, anyone who *says* that it is only a game has broken the cardinal rule of the game. Within the game, we pretend that we are in the fictional world we are imagining and in so doing we also pretend that the fictional world is complete. This pretense of completeness is maintained by imaginatively fleshing out any feature of the fictional world which is left underdetermined by the author that would otherwise threaten to break this first rule of the pretense, viz. that it is always uncouth to *say* what all collaborators should already *know*: “It’s just a game.” Consequently, we need not imagine every detail of a fictional world in order to treat it as complete. We need only imagine those details which we find necessary for participation in the game.

While this mentalist Waltonian stance is capable of explaining the intuition that fictional worlds are complete, it is *not* capable of explaining the other side of the bifurcation: the author’s prescription to us to imagine certain things and our critical restriction to speak only of these certain things. In Section 3.2 we determined that the prop of a work of fiction, the real-world object or event which fixes the ways in which we apply the principles of generation for our pretense, is an *impression* of the actual world. What allows Timmy and Tommy to collaborate in a game of make-believe was the fact that there were tree stumps to which both boys have access. When a Petersburg-born Russian reads *Crime and Punishment*, she will certainly have a much different impression of Petersburg upon which to build her pretense than I will. Though she might say that the difference between her impression and mine is trivial where *Crime and*

*Punishment* is concerned, it seems as if Dostoevsky must have assumed that there is some similarity between my impression of the actual world and his impression of the actual world, despite the fact that no one has access to anyone else's impressions. So if ficta and fictional worlds are entirely mental, then when I speak about Holmes, when you speak about Holmes and when Conan Doyle speaks about Holmes, all three of us refer to *different* objects, and these objects are *private* to each of us. Though Walton proposes that props prescribe, there is still no mechanism which can explain how it is that the principles of generation for *Crime and Punishment* do not produce a unique and incommensurate pretense for each person. We conclude that this naïve form of mentalism is still not enough to secure either an adequate theory of pretense or an adequate theory of fiction.

#### **4.2 Collingwood's Theory of Art**

The contemporary philosopher is in the comfortable position of having an enormous library of previous theories from which to borrow in times of either need or laziness. This contemporary philosopher has found R. G. Collingwood's theory of art, established in *The Principles of Art*,<sup>24</sup> to be adequate to the task of addressing both sides of the Bifurcation, which our naïve mentalism could not do. In accomplishing this task, we will first lay out Collingwood's theory of art – a complicated theory of sensation, imagination, expression and language; then, in the following section, this theory will be applied to our specific case of canonical works of fiction.

Collingwood's theory of art maintains that all artworks are expressions of emotions felt by the artist (author). Among these emotions, the simple and raw ones are the psychical emotions; whereas, the more complex ones are conscious and intellection emotions. What is

---

<sup>24</sup> R. G. Collingwood, *The Principles of Art*, (London, Oxford, New York: Oxford University Press, 1938). Hereafter cited in text as *Principles*.

peculiar about Collingwood's theory is that emotions are characterized so broadly that any act of language is an expression of emotion. In fact, the identification of expression with language will be crucial to this sophisticated form of mentalism, for it will solve the problem of externality to which naïve mentalism fell prey. Another crucial feature of this mentalism is that it explains how the impression of the actual world can be used as a prop for a fictional world: on Collingwood's theory, there is no qualitative distinction between memory and imagination. The only difference is that one is constrained to construction based on *sensa* and the other is not so constrained. Let us therefore begin with sensation.

#### **4.2.1 Sensations and Emotions : The Constituents of Human Experience**

According to Collingwood, human sensory experience takes the form of a union of sense-data with an emotional charge (conceived analogously to the primary features of a sub-atomic particle: mass and electric charge). The various color tactics that advertising companies implement nowadays are evidence that even on the simplest sensory level, a sense-datum (or *sensum*, as Collingwood calls them), as experienced, is primordially inextricable from its "emotional charge". Collingwood recognizes that many persons nowadays have 'sterilized' their sensory experience in an effort to tune out the emotional charge:

In persons who are likely to read [*The Principles of Art*], the habit of sterilizing sense has probably become so ingrained that a reader who tries to go behind it will find it very hard to overcome the resistance which hampers him at every move in his inquiry. In so far as he succeeds in recognizing what really happens in himself, I believe he will find that every *sensum* presents itself to him bearing a particular emotional charge, and that sensation and emotion, thus related, are twin elements in every experience of feeling. In children this is clearer than in adults, because they have not yet been educated into the conventions of the society into which they have been born; in artists clearer than in other adults, because in order to be artists they must train themselves in that particular to resist these conventions (*Principles* 163).

The most basic forms of emotions are the passive emotions, which are later either perpetuated or denied by the consciousness which experiences them. On Collingwood's theory of art, it is the task of imagination to examine and explore these sensations and emotions by holding onto them



for longer than the passive experiential phenomenon which initially brought about the sensations and emotions would normally allow. Thus, there is something active and *free* at work in the act of imagination (*Principles* 197).<sup>25</sup> This freedom lies in the concept of attention: when our conscious mind attends to a feature of experience (whether sensory or emotive), it chooses to preserve that feature for longer than the short period of time provided to us by our passive sensory and emotive experiences – it divides all these experiences into (a) the unattended: those which vanish after the duration of our experience of them lapses; and (b) the attended: those which remain when the experience is finished (*Principles* 203-6). Any act of recollection is an example of this phenomenon, for the concept of memory is intricately bound to the concept of imagination. The more intensely we attend to an experience, the greater the clarity of that experience. Peripheral vision is a readily available example of this fact.

As part of the extension of experience *simpliciter*, imagination does not provide us with anything universal. All of our imagined sensations and emotions are singular, particular and experienced only once – and this Collingwood takes to be an obvious feature of human experience. In a trivial sense, the racecar that you imagine cannot possibly be identical to the racecar that I imagine because you and I are different persons. The dissimilarity begins with non-identity, but it goes beyond the mere fact that you and I are not identical. Because of this non-identity, the only means by which you and I can even compare our sensations and emotions is by somehow expressing them to each other and deciding each for ourselves, based on this expression, how similar your sensations and emotions are to mine. Thus, only if I tell you that my racecar is a white and red formula 1, you will be able to determine whether there is some

---

<sup>25</sup> I interpret Collingwood to use the word ‘freedom’ somewhat phenomenally and thus independently of metaphysical implications. If simply refers to the experience of choice, regardless of the debate between libertarians and determinists, then we may take Collingwood to be accounting for the *experience of activity* versus the *experience of passivity*. And the concept of freedom accounts for this distinction.

similarity between our imaginings – at least concerning the criteria of color and shape. Despite these similarities, there are many possible *dissimilarities* between our imaginings which have not even been spoken. It is these unspoken yet highly likely dissimilarities which secure the non-universality of emotions, and part of an author's charge is to expose some of these differences by generating contention and disagreement about meaning via explicitly non-abstracted (and therefore non-universalized) expression. "The poet... , in proportion as he understands his business, gets as far away as possible from merely labeling his emotions as instances of this or that general kind, and takes enormous pains to individualize them by expressing them in terms which reveal their difference from any other emotion of the same sort" (*Principles* 113). The poet does not express his emotions such that the audience will feel the *same* emotion, for this is impossible; rather, he expresses his emotions such that each member of the audience will express to herself an emotion similar to the one the author felt.

#### **4.2.2 Conscious Emotions: Expression of Emotion or Expression as Emotion?**

One of the essential features of art is the expression of emotions. But what does 'expression' mean? Collingwood goes to great pains to make sure that the reader does not interpret expression as arousal – the first of three parts of *The Principles of Art* is dedicated to this precaution. An activity which arouses human emotions, he observes, assumes a purpose for the arousal: it is a means-ends activity. The purpose for arousal is what separates different forms of emotional arousal from each other. Commercials arouse emotions whose purpose is to effect an action concerning their products just as politicians arouse emotions whose purpose is to effect an action concerning votes. If the purported artist arouses emotions for their own sake, she has produced entertainment; if she arouses them to bring the observer to activity (as in oratory rabble-rousing), then she has produced what Collingwood calls 'magic'. On Collingwood's

theory, these are all crafts – not arts – because they assume that the activity of art is a functional means-ends activity.<sup>26</sup>

Before we further examine what it means to express an emotion or sensation, we must determine what it is that emotions and sensations are. Collingwood's ontology of emotions and sensations is determined by the psychological layers of the agent experiencing the emotions and sensations.<sup>27</sup> The first set of passive experiential phenomena is sensory experience, which Collingwood calls 'sensa' as distinguished from the activity-connoting form 'sensation'. This level is nothing more than the sense-data about which Russell spoke when he described his desk. Pure sensation requires absolutely no cognitive activity: it can be argued that even computers have this capacity.

The presence of a psyche, however, leaves emotive impressions upon the agent. Bulls charge with rage at red cloth and cats react with fear at certain shapes. This is what Collingwood calls emotional charge – the second set of passive experiential phenomena. Whether or not this emotional charge is *imposed upon* the sensa by the psyche (emotive realism) or is *detected in* the sensa by the psyche (emotive anti-realism) is irrelevant, because the phenomenon of emotional charge is both passive and private: neither can this question be answered, nor does answering it have any effect on Collingwood's system. Let us just say that these psychological emotions are concurrent with sensations and exclusive to agents who possess psyches.

Sense-data and psychological emotions exhaust the experiential limits of a pre-conscious agent, yet these two sets of experiential phenomena (sense-data and psychological emotions)

---

<sup>26</sup> Collingwood denies that art is a means-ends activity, but he does not explicitly state what other sorts of activities there are. But if we ask the question "what is the purpose of art?" then we have already assumed that it is a means-ends activity. In positing the existence of human activities which do not have a means-ends structure, he is implicitly concurring with Wittgenstein on the matter: This is simply what we do. Collingwood is, of course, careful not point out that although art is not *essentially* a means-ends activity, it can still serve this function secondarily.

<sup>27</sup> Collingwood does not define the word 'psyche', though there are myriad available examples of animals whose cognitive activity is limited to what Collingwood calls psychological emotions (aggression, fear, pleasure, lust, etc.).

comprise, in one sense, the totality of the passive content of human experience. In fact, these two sets of passive phenomena are already enough for the existence of simple languages, such as animal communication. And so for Collingwood, animal communication is the expression of psychical emotions via behavioral signs which are *common* to the animals doing the communicating.<sup>28</sup> Cats wink at and meow at each other because these are their natural modes of *expression*, where to express an emotion is to behave a certain way upon the experience of an emotion. The natural mode of expression for any creature is nothing more than the set of actions that accompany the experience of the emotion and/or sensation. For Collingwood, communication comes about by the simple fact that beings with psyches react to stimuli and recognize similar reactions in others: two animals of the same species may communicate because they naturally express emotions in the same way. This is the genesis of expression, which, for Collingwood, is the original form of language. For cats, a hiss expresses aggression, a wink expresses non-aggression, a purr pleasure, a growl displeasure, etc. These unconscious expressions can be thought of as instinctive rituals accompanying the emotions they express, for an unconscious animal cannot but express emotions it expresses in the ways that it does.

“But,” an attentive reader might complain, “human experience is much richer than the mere psychical emotions: we love, we despair, we ire, we hate.” Collingwood is more than sympathetic with such a reader; for the psychical emotions, thus far described, are not the subject of art. Pure psychical emotions like aggression and fear are too raw and animal for their expression to be termed ‘art’, for art is delicate and subtle rather than brutish and coarse. The emotions expressed in art require effort to express, and they take time to understand (re-express). Although expressions of fear and aggression are often component parts of an artwork, these

---

<sup>28</sup> Wittgenstein expressed this fact when he said that animals and human beings share a “form of life”.

emotions (as well as their expressions) are not sufficiently organized to *constitute* the expression of an artwork. For art, as a distinctively human enterprise, must be a conscious act.

The emotions our attentive reader has mentioned, on the other hand, *are* complex enough to be the subject of art, for these emotions are conscious emotions. Intuitively, there is a respect in which conscious emotions are passive, for sometimes one cannot help but feel angry or in despair. But there is also a respect in which they are active, as any adult should know by now: anger, for example, is an emotion that must often be neutralized for the sake of civility.

A conscious agent bears the capacity for attention which, as mentioned above, preserves a passive experience and converts it into an active experience: attention lies at the heart of imagination. A conscious agent who also uses language to express – via that agent’s natural mode of communication (e.g. voice, pictorial design, body language, etc.) – the *sensa* and psychical emotions she experiences is now in what Collingwood calls a “double situation”.

By definition, a conscious agent is aware of itself, as is implied by the freedom of choice in the act of attention. And self-awareness entails that any act of language from one conscious being to another is an act of expression both to itself and to that other. E. g. the same phrase in the mind often means something different once it receives the subtle nuances of both vocal and body language, and this is why persons who are surprised at the audience’s reception of a thought freshly spoken often find themselves in embarrassment. Provided that there is such a difference, every act of speech is an act of expression: it is just as new to the speaker as to the listener. The expression expresses to the speaker and to the listener at the same time; it does not merely repeat.

So the double situation is the fact that any language user is always both a speaker and a listener, regardless of whose vocal cords are vibrating. The act of speaking is essentially the

expression of an emotion to another conscious agent, so if one is alone in the room, one must be expressing to oneself. It is for this reason that the act of understanding is the act of expressing. When I understand a poem, I feel the emotions that it expresses, but the means by which I feel those emotions is to express them to myself using the words of the poem. Thus, all persons who experience an artistic expression are collaborators in that expression, because they all express emotions, though they do borrow a medium of expression. On Collingwood's account, then, the work of art, which is the simultaneous act of expression and feeling or, as Collingwood terms it, a "total imaginative experience," is not essentially connected to the physical bodies which we typically refer to when we speak the word "art"; rather, the artifacts produced by artists are merely tools used in the creation of art proper – which the artist intends to exist in the mind of the observer. Yet there are many objects not fashioned by an artist which still prove to be occasions for the existence of art in the mind of an observer. The artist's goal is not only to produce art in the mind of an observer, but to do it *immediately and frequently*. But while Collingwood himself is a mentalist about art, it seems to me that often we cannot experience art with *clarity* unless the physical body is present. This fact is most apparent in imaginary sensations: the fine details of Michaelangelo's *David* are much easier to imagine when the sculpture is in front of you than at any other time.

This double situation provides conscious language users with a new capacity: the relationality which is defining of our self-awareness has afforded us the ability to organize and modify the various emotions which we express. For after we express to ourselves an emotion (or after some other person does so), first we recall both the emotion and the expression imaginatively, then we decide whether the one has adequately expressed the other, and finally we adjust the expression accordingly. Whenever we correct another person's expression, we apply

our conscious organization upon her expression. This is how we recognize emotions in others, for Collingwood: by re-expressing their emotions to ourselves. Consequently, on Collingwood's phenomenology of emotions, the artist both experiences conscious emotions and expresses them at the same time. So the act of expressing an emotion is not only the act of clarifying and distinguishing it from other emotions, but it is also constitutive of that emotion:

[T]he expression of emotion is not, as it were, a dress made to fit an emotion already existing, but is an activity without which the experience of that emotion cannot exist. Take away the language, and you take away what is expressed; there is nothing left but crude feeling at the merely psychical level (*Principles* 244).

In fact, because every emotion is different and because we cannot recognize a conscious emotion until it is organized, no conscious emotion is felt *unless* it is expressed.

The emotions expressed by an artist through a work of art are not always expressed by the artist alone, for there are often collaborators: actors, dancers, and violinists to name a few. But Collingwood does not end artistic collaboration with these traditional collaborators. Rather, the entire audience, for Collingwood, collaborates in the artwork. This is because each person in the audience finds herself in the double situation which demands simultaneous expression and experiencing. Thus, members of the audience express the content of a *Macbeth* to themselves using the materials provided by Shakespeare and the actors, thereby becoming themselves collaborators in the artwork that is *Macbeth*. This is why no play is *performed* without an audience: otherwise it is only *practiced*.<sup>29</sup>

So let us consider an obvious example of such a conscious emotion: love. In one sense love is composed of a combination of psychical emotions (which are likely to include pleasure, aggression, non-aggression, fear, lust or *eros*, and more). But, due to the double situation of the

---

<sup>29</sup> While the actors themselves count as spectators and collaborators, they do not count as an audience because their act of expression is memorized and rehearsed. They do not have the fresh, almost naïve eyes of an audience. This is the crucial contribution of an audience.

conscious agent, in another sense it is composed of a structure consciously imposed by the agent to organize the psychical emotions into a coherent human emotion. It is unlikely that I understand what it means to love my wife if I express aggression, pleasure or lust toward her in a disorganized fashion. In fact, it is more likely that I will be considered either insane or emotionally immature.

Poets have expressed to us many times in their poems the mysteries and nuances of love, jealousy, courage and other conscious emotions, and often the reader learns something new about these emotions from each poem. Because that which is complex must be built out of that which is simple, the raw and crude psychical emotions are an obvious choice for the building blocks of conscious emotions. For this reason, Collingwood likens the psychical emotions to the concept of Aristotelian matter: they are the stuff out of which all other emotions are fashioned. Similarly, conscious emotions are likened to the concept of Aristotelian form: they are the organization and structure of psychical emotions. Therefore, the differences in form and distribution of psychical emotions within a conscious emotion are what distinguish each conscious emotion from the others.

#### **4.2.3 Intellectual Experience: The Formalization of an Expressive Language**

The final set of experiential phenomena is the intellectual emotions. These emotions may only be felt and expressed once the type of language discussed until this point – conscious expressive language – is fitted and formalized by means of grammar and logic, that it may serve the organizational needs of the intellect. But because no two emotions are alike, grammar and logic cannot be used to express everything that there is to say – some emotions just won't fit the mold. Collingwood expresses the violence done to language by grammar in his depiction of grammarians as butchers (*Principles* 259). Grammarians assume that expressive categories are



distinct (*Principles* 254-9), just as logicians assume that thought must be expressed according to certain forms. Imprecise though they may at times be, the goal of introducing grammar and logic is to “make language into a perfect vehicle for the expression of thought” (*Principles* 259), and we have clearly had much success.

The intellectual emotions expressed through formalized language arise out of situations “which could not generate them unless [they] were intellectually apprehended.” This should not suggest that intellectual emotions are expressed only in formalized language. Rather, formalization *permits* expression of these emotions, and this expression often occurs in tandem with conscious emotions:

[T]he poet converts human experience into poetry not by expurgating it, cutting out the intellectual elements and preserving the emotional, and then expressing this residue; but by fusing thought itself into emotion: thinking in a certain way and then expressing how it feels to think in that way. Thus Dante has fused the Thomistic philosophy into a poem expressing what it feels like to be a Thomist. Shelley, when he made the earth say, ‘I spin beneath my pyramid of night’, expressed what it feels like to be a Copernican (*Principles* 295).

We now have all the materials with which to formulate a Collingwoodian definition of art. *Art is an activity: it is the simultaneous collaborative expression and experiencing of a complex sensory and emotive imagining with unique imaginings contributed by each member of the collaborating community.*

### **4.3 Applying Collingwood’s Theory**

Now that we have laid out the relevant features of Collingwood’s theory of art, we may equip it with the vocabulary we have already established prior to laying out Collingwood’s theory. This translation is generally a smooth translation, for Collingwood’s theory is a very suitable match for our own theory.

What is, for Collingwood, a total imaginative experience is, for Walton, an act of pretense or make-believe. When we say that for Collingwood, ficta are mentalia, we mean that

they consist of an active imaginative rendering whose *matter* is the raw and passively experienced mentalia – emotions and sensations – and whose *form* is the conscious and intellectual organization of these mentalia into a coherent imaginative whole. A pretense is a complex web of emotions, and because emotions are thoughts for Collingwood, a pretense is also a complex web of thoughts. But because no set of emotions and sensations is the same for any given person, there are as many fictional worlds for any given text as there are readers. If an expression is a purely imaginative act (and therefore private), then the expression and the emotion are not distinct from each other. Both expression and emotion are private to the individual expression and emoting, and, as such, they are indistinguishable from each other.

If, on the other hand, the expression is imposed onto an artifact, such as ink on bound paper, then the expression is distinct from the emotion in the following way: there is only one expression, but there many are associated emotions – specifically, one for each reader. This is participation in the creation of a fictional world, because each reader uses the author's words to express the fictional world to herself. It is to the extent that the reader is expected to participate in the creation of the fictional world that the reader is a *collaborator* in the act of art that produces the fictional world. This entails that whenever we refer to 'the' fictional world of a text, we are actually referring to a very large set of fictional worlds. As David Lewis proposed, there are myriad fictional worlds to which the stories about Sherlock Holmes refers, but the Lewisian claim that the set is infinite (or near infinite) because variations on the worlds are infinite (or near infinite) is a distinctly realist claim. A Collingwoodian mentalist ought to maintain that the stories written by Conan Doyle refer to as many Holmes worlds as there are collaborators in the act of imagining the world of Sherlock Holmes (and perhaps more, since some collaborators may imagine multiple worlds).

A distinction should be made between the infinite set of Holmes worlds which *may* be imagined and the finite set of Holmes worlds which *are* imagined. Strictly speaking, the Collingwoodian mentalist must maintain that the only Holmes worlds which exist are the worlds which have been imagined by real persons – all others are merely hypothetical Holmes worlds which *do not exist in any way* until they are imagined.<sup>30</sup> So how do these hypothetical Holmes worlds come to exist? They come to exist only when a person consciously attends to them. I alone can create as many Sherlock Holmes worlds as I like, though it is unlikely that I will spend the time necessary for the imaginative construction of these worlds. Based on these considerations, we can say that while the text of *Crime and Punishment* is an elaborate expression of the events of an imaginary world, the specific sentences written by Dostoevsky in the book itself are expressions of *only* the features of this imaginary world to which Dostoevsky wanted to draw our attention. Everything else he has left up to us to attend to (if we choose).

For Collingwood, intellectual language is an attempt to structure and formalize a smooth and organic layering of the expressions of a rich manifold of emotions and sensations. Because this task is impossible to complete in written language (or any language, for that matter), vagueness is unavoidable. Obviously, this is impossible because a dense world would require infinite expression in order for us to achieve the sort of *objective* descriptive precision which is possible in the actual world. But there is another reason that a complete formal transformation of written language is impossible. Written language lacks the capacity to express many of the sensations and emotions which are part of our experience of the actual world; however, written language can be used in tandem with other modes of expression in order to capture a broader

---

<sup>30</sup> Minimally, these hypothetical fictional worlds *do* already exist as a vague set of pretenses by virtue of the fact that I am imagining them right now. Each of these worlds, however, is virtually free of content. An alternate treatment of this phenomenon is that I am pretending to imagine an infinite set of Holmes worlds. The act of imagination is itself only a pretense because I'm not really imagining anything having to do with Sherlock Holmes; rather, I am only speaking as if I am.

spectrum of emotions. In Collingwood's colorful parlance,

if a civilization loses all power of expression except through voice, and then asserts that the voice is the best expressive medium, it is simply saying that it knows of nothing in itself that is worth expressing except what can be thus expressed; and that is tautology, for it merely means 'what we ... do not know we do not know', except so far as it suggests the addition: 'and we do not wish to find out' (*Principles* 246).

According to Collingwood's theory of art, all emotions are only experienced insofar as they are expressed, so if an emotion *cannot* be expressed in written language, then it *will* not be expressed unless an alternate medium of expression is used. Therefore, all texts are constrained by the emotional and sensory carrying capacity of the language in which they are written. And formal language has a smaller carrying capacity than a combination of formal and non-formal language.

If we accept Sorensen's suggestion to treat fiction sorites paradoxes as a consequence of semantic vagueness concerning an object (*ficta*) which is not vague, then the Collingwoodian notion that there is no qualitative distinction between our conscious experience of the actual world (about which standard sorites paradoxes are concerned) and our experience with fictional worlds (about which fiction sorites paradoxes are concerned) neatly fits our treatment of the fiction sorites paradoxes. For Walton, pretense is partially constitutive of the world in which we live, because the practice is simply inescapable even in everyday human affairs;<sup>31</sup> for Collingwood, our experiences of both kinds of worlds are located in the imagination: imagination is *completely* constitutive of both, though Collingwood would not describe memories as pretenses.<sup>32</sup> The similarity between the word 'noonish' and the name 'Sherlock Holmes' is that both words explicitly denote imprecision. The difference, however, is that the

---

<sup>31</sup> Walton holds this to be true by virtue of the depth and ubiquity of pretense-dependent art. However, there are other, more mundane examples of pretenses at work in a normal adult life. Currency, for example, is an act of make-believe: we pretend fiat paper money has trading value because everyone else also participates in the same pretense. The practice of 'workplace professionalism' is also an act of make-believe: we pretend that we do not have any personal involvement with the persons with whom we interact in the workplace. And so on.

<sup>32</sup> For Collingwood, the imaginative process which creates memories about *sensa* is constrained by the actual *experience* of *sensa*; whereas, the imaginative process which creates pretenses is only constrained by the pretense-governing rules which the agent has self-imposed.

one is caused by a strictly numerical imprecision, while the other is caused by a much broader imprecision – an imaginary imprecision, out of which numerical imprecisions galore can be gleaned and used to construct fiction sorites arguments. This imaginary imprecision is a feature of the character of Sherlock Holmes because no expression can possibly determine Sherlock Holmes as a dense, completed person – imaginary augmentation is necessary to complete any feature of Sherlock Holmes to which we attend and in which we want to find dense determination. In short, there is nothing about my imaginary Holmes which prevents me from giving him all the properties he needs to be as fully and densely determined as an actual person. What prevents me from doing so is that such a task could never be completed. So we see that semantic vagueness is surely the culprit in both cases: in the standard case, ‘noonish’ is an imprecise term which can be precisified in any specific usage (15 minutes after noon may or may not be noonish, depending on the size of one’s time cushion); similarly in the fiction case, ‘Sherlock Holmes’ is an imprecise term which can be precisified in any specific usage (*my* Sherlock Holmes has 154,511 hairs). These clarifications should abate Sorensen’s objection that anti-realist explanations of the sorites paradox tend to be *ad hoc*.

The Bifurcation in our speech seems less severe now because the *reference* of ‘Sherlock Holmes’ in any sentence about him is not completely determinate; it is only *spoken of as if it were* determinate, because any heretofore unspecified feature of Holmes that you care to consider is capable of being determined by your own imagination. But the bifurcation it is still present and still relevant, because the within/without distinction is still essential to the business of fiction. To speak of a fictional world from without is, for the most part at least, to assume ignorance of the experiences of imaginative collaborators – for these are private. But let us not forget that *all* emotions and sensations are private. It is not the privacy of the emotions that

causes literary criticism to ignore the experiences of audience collaborators. It is the privacy of the *contribution of the audience collaborators* which prevents literary criticism from considering any expressions but those written by the original author. Interpretations and variations of theatre, for example, are many, so critics must keep in mind what remains the same when comparing such variations. Literary criticism, then, focuses on the expressions of the author in an attempt to further clarify the structure of the emotions and sensations evoked by these expressions. So while literary criticism does ignore the experiences of collaborators, it is also an act of artistic collaboration, which is why poetry is often read in tandem with criticism. In effect, the literary critic is attempting to collaborate in a pretense whose principles of generation cannot change, for a text does not change. Introduction of no new principles means that anyone who intends to collaborate in such a pretense must explicitly demarcate a boundary between imaginings which explicitly or implicitly adhere to the principles given to us by the author and the imaginings which either break or are not covered by the principles given to us by the author. Sherlock Holmes may have 154,511 hairs on his head, but Conan Doyle did not imply this number anywhere in his stories, so if I believe it, it is not according to any principle of generation that I do so. Consequently, I cannot expect anyone else to accept that this is the number of hairs on his head. This takes care of the critical (disengaged) side of the Bifurcation.

Collingwood's account of sensation is relevant to our project because of the similarity between memory and objects of imagination. While *sensa* come about through involuntary means, those sensations which we commit to memory are in part a voluntary set: that set of sensations to which we attend. For Collingwood, there is no qualitative difference between memory and imagination. The only difference is quantitative: depth and degree of precision. Because the only constitutive difference between memory and constructive (i.e. fictional)

imagination is of a quantitative nature, the truth-makers for fictional sentences are equally similar to the truth-makers for non-fictional sentences. When I say ‘Sherlock Holmes plays the violin’ and ‘Barack Obama plays basketball’, each of these sentences is only spoken *as true* by virtue of the fact that they are true in their respective imaginative worlds. Of course, the difference between these sentences is that one may be checked against the physical world and the other may not. But it is not the act of playing basketball which is “out there” in the actual world which makes the sentence ‘Barack Obama plays basketball’ spoken by me as true. If the fact, independent of human thought, that Obama plays basketball is what distinguishes a sentence spoke as true from a sentence spoken as false, then no one could possibly speak the sentence as false. The reason that we all agree that Barack Obama plays basketball is *not* that the sentence is true independent of human cognition. This may be, but, as Kant often reminds us, we have no access to such a fact. We have only access to our experience of the physical world, and the partial passivity of this experience is what secures agreement concerning Obama’s basketball habits. This sentence is *not* widely thought to be true by virtue of *mere coincidence* – we don’t all just incidentally imagine the same reality. Rather, reality, as experienced, involves acceptance of the proposition that Obama plays basketball. We have, here, a distinction between a sentence spoken *as true* by the speaker, and a sentence which *is true* independent of the speaker. The first is relevant to a fictional mentalism, and the second is not. The similarity between the truth-makers of fictional and non-fictional sentences is the means by which Collingwood can account for the fact that we apparently speak fictional sentences as if they were spoken about the actual world. The similarity extends beyond mere seeming, for our accounts of the actual world originate in an act of imagination just as our accounts of imaginary worlds do. And this takes care of the fictional (engaged) side of the Bifurcation.

It should be stressed that Collingwood was an aesthetic idealist, maintaining that “aesthetic experience ... is wholly and entirely imaginative” (*Principles* 306). But this experience does not arise out of thin air, for emotions require expression in order to be felt. As mentioned previously, the purpose of the physical work is to facilitate the observer’s own imaginative expression. In most cases, a physical work is a necessary condition for the existence of art: a story must be written or told, an image shown, a song played, etc. But despite the frequent dependence of an artist upon her artifact, Collingwood believes that art is not essentially physical, a fact evident in our ability to keep a tune in mind all day after having heard it only once. In the case of traditional works of fiction, the text or the telling of the story is a necessary condition for the existence of the art. The physical work is a necessary condition for the author’s ability to provide others with the expression of his fictional world, but the fictional world may still be experienced without access to the text. In fact, we can imagine simple fictions to ourselves without any physical activity at all, thus demonstrating that the physical text is not essential to the work of fiction.

Yet there is still something artificial about the Collingwoodian explanation of the bifurcation in our speech about fictional worlds. Except in the special case in which a person imaginatively expresses to himself, the physical body is *absolutely* necessary to *all* acts of expression. Assuming that the artist intended to share her expression, how could we recognize an expression if there were no physical sign? This should suggest to us that art *not* is purely mental, as Collingwood thought. Emotions are purely mental,<sup>33</sup> but any expression which ventures beyond the mind of the person expressing is purely physical. When I speak, my words are

---

<sup>33</sup> An emotion – as felt – is an experience, and is thus purely mental. Naturally, there are many physical processes that accompany the experience of an emotion (blushes, hormonal releases, laughter, neuron firing, etc.), but these are all physical processes distinct (though associated with) the mental experience. Whether mental experience can be reduced to physical events is not relevant to our project; we simply take the distinction between the two kinds of events (mental and physical) to be obvious, even if an explanation for the distinction is not.



vibrations in the air, and when I draw, my pictures are marks on a paper. If I do not speak, then no one hears my words and my emotion is left (communally) unexpressed. Though Collingwood seems to believe that physical expression is not necessary to the creation of art (and spectators are, therefore, merely eavesdroppers), the genesis of communication as Collingwood describes it seems to be in tension with such a firmly mentalist position. For animals, an expression is a physical manifestation of the emotion imaginatively experienced, and this manifestation cannot be helped:

[The most primitive kind of expression, psychical expression] consists in the doing of involuntary and perhaps wholly unconscious bodily acts, related in a peculiar way to the emotions they are said to express. Thus, certain distortions of the face express pain; a slackening of muscles and a cold pallor of the skin express fear; and so forth (*Principles* 229).

Every kind and shade of emotion which occurs at the purely psychical level of experience has its counterpart in some change of muscular or circulatory or glandular system which ... expresses it (*Principles* 230).

But it is only the *psychical* expressions and emotions which are involuntary. *Conscious* emotions are felt voluntarily, for they must be organized before they are felt. This means that their expression is also voluntary, and Collingwood's position that art is completely mental remains (momentarily) intact.

We still seem to have the problem that naïve mentalism had: given that art is a completely mental affair, there is no security in Dostoevsky's assumption that the fiction I imagine when I read *Crime and Punishment* will be similar enough to the fiction that *he* imagines that we can speak of them loosely as the same fiction. Collingwood himself believed that the artist's act of expression was only *incidentally* expressed in physical terms and that, in principle, the expression may already be completed in the mind of the artist. But this is not always how art is created. Often an artist will start a project, let it sit for a moment and come back to it with fresh eyes. The act of expression, in such an instance, occurs in response to that

which is already partially expressed. In this sense, we may say that there is a necessary externalization involved in producing art.

In contradistinction to Collingwood's belief that an expression might be private, it will be useful for us to recall one of his contemporaries: Ludwig Wittgenstein. For Collingwood, expression *is* language, this identity allows us to apply Wittgenstein's Private Language Argument to *any* act of expression.<sup>34</sup> To put this argument loosely in Collingwood's terms: there is no emotion that cannot be expressed somehow, and because we are all capable of experiencing these emotions, we are all capable of understanding their expressions when we see them. Therefore, there is no expression which is in principle private; there are only expressions which remain private by virtue of having never been spoken. But even in this last case, the chances are that some artist will eventually experience an emotion similar enough that his expression renders the unspoken expression obsolete. We conclude that Dostoevsky's assumption that I will imagine a pretense similar enough to the one he imagined to describe them both as the fictional world of *Crime and Punishment* is safe. And the safety of this assumption extends beyond the fact that there is no in principle private language. The safety of this assumption comes from that fact that human beings use language to express their impressions of the *actual* world on a regular basis. Dostoevsky knew what knowledge he could assume his readers had because he spoke to people about their impressions of the actual world on a regular basis.

It may be useful to think of the text of a work of fiction as analogous to the foundation of a building. The foundation is often composed of an entirely different material from the rest of the building, and it is poured as a single solid whole; whereas the building itself is constructed progressively and in parts. In our case, the text of *Crime and Punishment* is the foundation of the

---

<sup>34</sup> *Philosophical Investigations*, tr. G. E. M. Anscombe, 3<sup>rd</sup> ed., (Oxford: Blackwell Publishing, 2001), ~ §256-§271.

work of art. It comes as a single unit, for we cannot add or remove anything from a completed work of art without turning it into a different work of art. This solid edifice gives us a foothold for constructing an imaginary world in which Raskolnikov, Svidrigailov, Porfiry Petrovich, Sofia, and the rest come alive. But, to break our building analogy, the fiction is never completed. It always bears the potential of completion, because most texts are sufficiently robust that a completed fictional world *may* be imagined (if a powerful enough mind cared to do so), but it never *will* be. Rather, we *attend* to the parts of the world which interest us, and we build these parts up in order to express to ourselves the various emotions that are latent in the fictional worlds for which Dostoevsky poured the foundation.

The Bifurcation in our speech about fiction, then, is better explained as a distinction between the physical part of the work of fiction and the mental part: the body and the reasoned activity. Collingwood is correct that the mental aspect of art is what is really critical to the human mind – for this is the activity of art. Whatever physical activity an artwork might have, such as the moving light of a film or the actors in a play, this kind of activity is not considered to be essential to art – otherwise aerobics would be art. Art, being a human institution, is both a *body* and an *activity*; nevertheless, as in all human institutions, the body is conformed to the activity for which it is used. We don't care about ink-marks in a set of bound pages *per se*, and we don't care about paint splattered upon a canvas *per se*. If we did, all painted and inked objects would be art. No, what we care about is the expressions manifested in these works.

But let us not make the mistake of suggesting that fictional worlds exist *within* the text, for this would be analogous to the claim that thought exists within the body. In fact, thought exists within the *mind* – and whether the mind exists within the body is a question for a different essay. Similarly, we shall say that fictional worlds – which are ultimately imaginative thoughts –

exist within the mind, but just as human thought does not exist without a body, so *collaborative* fictional worlds also do not exist without body (and private fictional worlds do not exist without a body *in principle*). The text, however, is more than just a physical body. It is a physical body with a pattern imprinted upon it: it is an expression. Therefore, it is not the body itself which is of interest to the fictional ontologist; it is the expression embedded within the body, the pattern of characters. If we were to take the body as being relevant to the fictional world, then we would have to account for the fact that there are so many copies of *Crime and Punishment* existing in so many different forms. Yet the expression of *Crime and Punishment* is a unity (neglecting translations – of which my copy is one): there is only *one* canonical expression of *Crime and Punishment*, and this is the particular set of Russian sentences originally published by Dostoevsky. Thus, the canonical expression, as a pattern of thought impressed upon matter, is the only part of a fictional world which is necessarily material. As far as literary criticism is concerned, all other collaborative fictional expressions and emotions may be (and often are) completely mental. And yet the very act of literary criticism *is* collaboration in a pretense, an expression. So while Collingwood has *understated* the importance of the physical body, this understatement is forgivable because the importance of the physical body is easy to *overstate*.

Given this account of the text, how does Walton's concept of a prop fit into the mix? In one sense, the concept of a prop is not necessary to a work of fiction. Collingwood's account of the imagination is also an account of language. For Collingwood, all language is the expression of emotion – sometimes psychical, sometimes conscious, and sometimes intellectual. The act of expression is the physical manifestation which (in principle) necessarily accompanies the act of experiencing the emotion, and this act of expression functions as an act of communication because there are others who express their emotions in just the same way as we do. Language

comes about because we recognize an act of expression as such, and we can use another's act of expression to express similar emotions to ourselves. Consequently, a text need not be prescriptive, despite Walton's belief that it is. Though texts are written to be read, the emotions expressed in the text would not even be felt by the author if he did not write the text! So even if the author does not intend the work to prompt an act of make-believe, it will still do so for any reader who is *capable* of experiencing the emotions expressed therein (provided that the author's expression is complete and coherent).

In another sense, however, the prop is the sensory input which is retained by committing images to the memory via attention. Because memory-imagination exists prior to constructive-imagination, it is clearly the memory-imagination which provides the raw materials out of which constructive-imagination can build (just as Locke thought). Although in theatre, props are physical objects and events which are experienced, the actual existence of these props is not necessary to the *function* of a prop. What is necessary to the function of a prop is that physical objects and events are experienced *at some previous time*. A play certainly couldn't take place without props, but the presence of props is not essential to every art form. In the case of fiction, previous familiarity with props is all that is essential: we must have had some experiences and retained some memories out of which to build a pretense.

Our ontology of pretense, then, is as follows. A pretense is an imaginative experience and, as such, is a *thought*, a mental entity. A *shared* pretense, on the other hand, necessitates *physical* expressions – communications between persons who are involved in their own private pretenses. In communicating these expressions to each other, persons involved in private pretenses extend the pretense to the public domain. Literary criticism concerning a work of fiction such as *Crime and Punishment* is a community of expressions built around the text, which

holds a special position in the community: it is the master expression to which all other expressions (within the discourse of *Crime and Punishment*) must be subjected. The phenomenon of communication, in a Collingwoodian context, is collaborative mental act whose effects can be seen in the various expressions by which we communicate.<sup>35</sup> Consequently, despite the involvement of the material world in our pretenses, expressions, acts of make-believe, and fictional worlds are inter-subjective thoughts constituted by the collective imagined sensations and emotions of those participating in the pretense. But these collaborative pretenses have nothing in them which exists over and above the private imaginative experiences of each collaborator and the expressions shared by the collaborators. A collaborative pretense is an organized group of private pretenses, just as a conscious emotion is an organized group of psychical emotions. Therefore, our anti-realist ontology can be summed up: ficta and fictional worlds are mentalia; private fictional worlds are expressed externally in principle only; and shared fictional worlds are expressed externally of necessity, but the artifact is only a vehicle for the expression.

#### **4.4 The Mentalist Interpretation**

We now have a plausible resolution to our problem. Let us interpret. First, we observe that because ficta and fictional worlds are said to exist as mentalia, we no longer need the two layers of pretenses which Everett and Yablo prescribed to us. There is certainly a pretense involved in speaking about Sherlock Holmes, but we do not need to adopt a pretense in which the realist's ontology of ficta obtains. Rather, we maintain that a pretense is an act of imagination which is confined to the human mind, but which can be communicated via expression to other minds. The benefits of dropping the second layer of pretense are twofold: first, we have

---

<sup>35</sup> John McDowell defends a position very similar to this one in *World and Mind* (Cambridge, Mass.: Harvard University Press, 1994).

simplified our semantic mechanisms; second, and more importantly, we may now use our concept of fiction as a basic notion of pretense by which other (non-fictional) discourses may be reduced to fiction via the two-layer pretense theory. In any case, our interpretation of the existential class of sentences will be non-literal. ‘does not exist’ must be taken to mean ‘does not exist independent of human cognition’.

Because fictional worlds are imaginary, they take the same shape in our minds as the actual world does. That is, we imagine ourselves *experiencing* these fictional worlds. So fictions may be attributed all the same properties as real persons. This means that the only interpretation that must be done is to note which of the two contexts – the engaged context, in which we are imagining a *private* fictional world; or the disengaged context, in which we are collaborating in an act of imagination which is governed by a master expression, or a *public* the fictional world – is the appropriate context for the sentence. Interestingly, this distinction is already present in a literal interpretation of sentences (F<sub>1a</sub>), (F<sub>1b</sub>), (F<sub>2a</sub>), and (F<sub>2b</sub>) versus a literal interpretation of (F<sub>2c</sub>)! The first four sentences refer to Holmes, Raskolnikov and Porfiry Pretrovich as if they were persons; whereas, the fifth sentence refers to Raskolnikov as if he is a *character*. Consequently, the only interpretation that must be done to accommodate the Bifurcation is to note which context is intended (if the context is not already evident). We will simply borrow the fictional fictionalist notation of subscripts: Holmes<sub>1</sub> refers to the person, Holmes<sub>2</sub> refers to the character, and Holmes without subscript refers to both.

One of the greatest benefits to adopting Collingwood’s mentalism is that it entirely eliminates the dependence upon propositions which was fundamental to Walton’s conception of fictional worlds. For Collingwood, fictional worlds are no longer constituted by propositions, but by acts of imagination: they are constituted by emotions and sensations. This greatly simplifies

our semantic machinery. No longer do we need to concern ourselves with making sure that all our propositions are true, which was especially a problem for the fictional class of sentences – the engaged context. Rather, each sentence expresses an emotion, regardless of whether that sentence is true or false when spoken in the engaged context.

We also need not distinguish in which fictional world the character exists, because, as a mental entity, all characters exist in *this* world though they are all private to each person who imagines them. The complex semantic acrobatics which are meant to explain how a sentence about Sherlock Holmes could be true when Sherlock Holmes only exists within a pretense are no longer necessary because pretenses exist within the mind, so anyone who speaks the sentence speaks it as true of only himself. Incidentally, we have all chosen to agree that Sherlock Holmes plays the violin, because we all play the same make-believe game. So the truth-maker of the sentence ‘Sherlock Holmes plays the violin’ is different for every single instance of the sentence.

One might think that interpretation of comparative critical sentences is required because there is comparison across worlds. But in a Collingwoodian mentalism, this is not so. For Collingwood, our experience of the actual world is not qualitatively different from our experience of a fictional world, so as long as the comparison across worlds occurs in terms which are commensurate between these two worlds, there is no issue in taking the sentences literally. An example of worlds which are not commensurate would be attempting to compare, for example, the world of *Superman* with the world of *Lord of the Rings*. We cannot really say whether Superman is stronger than Sauron because we have no real bar for comparison. We might invent a fictional world in which the characters coexist, but this would still tell us nothing of the original comparison. So what makes two worlds commensurate is that the worlds have similar principles of generation. If there is sufficient difference between the principles of



generation of two worlds, comparison cannot be had. In *Lord of the Rings*, there are physical gods; whereas, in *Superman*, there are absurdly strong aliens. Each has a weakness (Superman has kryptonite and Sauron the One Ring), and each is invincible aside from that weakness. Because neither has fought a common foe besides normal human beings (who are not a challenge for either), it is impossible to answer the question “Who is stronger, Superman or Sauron?” This is how neither world shares the principles of generation of the other, and it is why we cannot say what would happen in a fight between Superman and Sauron.

And what of our analytic critical sentence? Our interpretation of ‘Raskolnikov is an archetypal character’ will depend entirely on how we define “character”. In addressing the Bifurcation on Collingwood’s theory, we concluded that pretenses are shared via communication concerning the pretense. Because our ability to communicate is limited and because we conform our communication concerning canonical works to the master-expression (the text), our critical discussion of fictional characters will be artificially limited to what was actually said by the author. This artificial limitation amounts to a cordoning off of those properties in our pretense which were expressed by the author from those properties which were not. A “character” has only those properties which were expressed (either implicitly or explicitly) by the author. Consequently, we may also interpret these sentences literally.

While fictional and critical sentences are interpreted literally in a Collingwoodian mentalism, we must never keep far from our minds the fact that references to Sherlock Holmes the person and to Sherlock Holmes the character both refer to different entities for each person who speaks their names. When I speak about Sherlock Holmes the person, I refer to my own mental construct and when you speak about Sherlock Holmes the character, you refer to your own mental construct. The difference between the person and the character is *not* that we refer to

the same entity when we speak about the character, but that the entities about which we speak when we refer to characters have the same *properties*. And this is not the case when we refer to the person. Thus, when we share a pretense, what we share is not the mental construct, but the *form* or *structure* of the mental construct.

#### Interpretation under Collingwoodian Mentalism

- (F<sub>1</sub>) (a''') Rodion Raskolnikov<sub>1</sub> is a murderer.  
(b''') Sherlock Holmes<sub>1</sub> plays the violin.
- (F<sub>2</sub>) (a''') Sherlock Holmes<sub>1</sub> is a more clever detective than Porfiry Petrovich<sub>1</sub>.  
(b''') Sherlock Holmes<sub>1</sub> is more famous than any real detective.  
(c''') Rodion Raskolnikov<sub>2</sub> is an archetypal character.
- (F<sub>3</sub>) (a''') Rodion Raskolnikov does not exist independent of human cognition.  
(b''') Sherlock Holmes does not exist independent of human cognition.

#### **4.5 Conclusion**

We began our essay by laying out the concepts which we take as central to any phenomenally faithful theory of fiction: (a) there is a problem of ontology in our discourse about fiction which demands a theory beyond mere common sense; (b) within our discourse about fiction, there is a sharp distinction which cannot be denied by anyone who partakes in the discourse: the engaged versus the disengaged context; (c) there are a series of intuitions which must be accommodated by our theory, most importantly, the intuitions that fictional worlds are (i) complete and (ii) created rather than discovered.

While our express goal was to devise a mentalist theory of ficta, the path to this goal took us through three alternate theories. The analytic tradition has provided us with two theories of ficta which are essentially versions of modal realism, augmented to fit the purposes of fiction. The purpose of examining these theories in an effort to produce an anti-realist theory is that, incidentally, one of them demonstrates the potential pitfalls that a theory of fiction might face; whereas, the other demonstrates the way in which a theory may be augmented in order to avoid

such pitfalls. Specifically, our examination of concrete realism demonstrated that, among other troubles, a purely realist theory cannot but fail to accommodate the Bifurcation in Speech. Our examination of abstract realism showed us (a) that a realist theory of fiction can avoid these pitfalls by incorporating an anti-realist concept of pretense and (b) that some form of anti-realism is *necessary* to the project of producing a theory of fiction. These discoveries suggested to us that our project of constructing a mentalist theory of fiction should begin with pretense and it should maintain a strictly anti-realist position.

After laying out the fundamental features of pretense theory as established by Walton and extended by Yablo, we found that a pretense theorist which claims to be able to explain fiction arrives at an ontological dilemma, either horn of which brings the pretense theorist to a Cartesian ontological conclusion concerning pretenses. We concluded in section 3.3.5 that the fictional fictionalist's dependence upon the ubiquity and reflexivity of pretenses ultimately entails an ontological claim similar to the *cogito*: 'I pretend, therefore I am,' which is revised to 'There is pretending.' And Collingwood's reduction of imagination to thought suggests that this similarity was not an illusion – the pretense *cogito* is a subset of the standard *cogito*, 'There is thinking.' While we make no claims about the relationship between mind and matter, the phenomenon of thought demonstrates to us unequivocally that if anything exists, thought exists. Therefore, an ontology of pretense was found to be entirely unavoidable, and the pretense theorist's attempt to reflexively (or circularly) explain fiction failed. In response to this failure of pretense theory to support a theory of fiction by itself, we naïvely took pretenses as mentalia (or thoughts) and adopted the fictional fictionalist's mechanisms for adopting the realist's pretense concerning fictional characters. But this naïve mentalism failed due to its inability to externalize pretenses,

for without some method of securing a reliable external pretense, we cannot accommodate the disengaged side of the Bifurcation.

We therefore rejected the reflexive pretense mechanism introduced by Yablo and Everett as a means of adopting the abstract realist's theory without committing ourselves to the abstract realist's ontology. In its place, we established a theory of expression which inextricably conjoined the experience of a pretense with the expression of that pretense. By reducing language to expression, we were able to borrow Wittgenstein's Private Language Argument to combat the failure of naïve mentalism – the inability to secure a reliable externalization of pretenses in order to accommodate the disengaged side of the Bifurcation. For Wittgenstein, despite the privacy of emotions and sensations, there can be no such thing as a human language which is private in principle. Similarly, though the only ontological commitments a Collingwoodian theory of fiction has are to human emotions and sensations, there is still a non-trivial sense in which we may communicate to each other concerning these emotions and sensations. This communication is the act of expression.

This form of externalization is not truly external to human cognition; rather, what is externalized is the human mind. For, as we all know, absent a sapient being who understands an expression (and thereby re-expresses the emotions originally expressed in the expression to himself), a text would be nothing more than a pile of paper and ink – so the thought never achieves independence from human cognition. The expressive habits of human beings are so similar to each other that there is a relevant sense in which a thought (emotion) can be shared by human beings – and this shared thought is captured in an expression. Consequently, we have an inter-subjective anti-realist ontology of ficta and fictional worlds on our hands. So our first

conclusion is that the anti-realist position endorsed herein amounts to a union of Waltonian phenomenology, and Collingwoodian theory.

Our second conclusion is that a comparison of the various interpretations of the three classes of sentences about fiction will make it obvious that the mentalist theory constructed herein is capable of providing the most faithful interpretations. The Collingwoodian theory, then, is the closest theory to common sense concerning discourse about ficta and fictional worlds. For the reader's convenience we will list the original sentences and all four interpretations below:

#### Types of Fictional Sentences to Be Interpreted

- (F<sub>1</sub>) (a) Rodion Raskolnikov is a murderer.
- (b) Sherlock Holmes plays the violin.
- (F<sub>2</sub>) (a) Sherlock Holmes is a more clever detective than Porfiry Petrovich.
- (b) Sherlock Holmes is more famous than any real detective.
- (c) Rodion Raskolnikov is an archetypal character.
- (F<sub>3</sub>) (a) Rodion Raskolnikov does not exist.
- (b) Sherlock Holmes does not exist.

#### Interpretation under Concrete Realism

- (F<sub>1</sub>) (a') In  $w_{CP}$ , Rodion Raskolnikov is a murderer.
- (b') In  $w_{SH}$ , Sherlock Holmes plays the violin.
- (F<sub>2</sub>) (a') 'Sherlock Holmes' functions as a more clever detective in  $w_{SH}$  than 'Porfiry Petrovich' does in  $w_{CP}$ .
- (b') In  $w_{SH}$  'Sherlock Holmes' functions as a more famous than any detective does in the actual world.
- (c') In  $w_{CP}$  'Rodion Raskolnikov' plays an archetypal role.
- (F<sub>3</sub>) (a') Rodion Raskolnikov does not *actually* exist.
- (b') Sherlock Holmes does not *actually* exist.

#### Interpretation under Abstract Realism

- (F<sub>1</sub>) (a'') Rodion Raskolnikov<sub>1</sub> is a murderer. [no truth value]
- (b'') Sherlock Holmes<sub>1</sub> plays the violin. [no truth value]
- (F<sub>2</sub>) (a'') Sherlock Holmes<sub>2</sub> encodes being-a-clever-detective to  $x$  degree of cleverness; whereas Porfiry Petrovich<sub>2</sub> encodes being-a-clever-detective to  $y$  degree of cleverness, and  $x > y$ .
- (b'') Sherlock Holmes<sub>2</sub> encodes being-a-famous-detective to  $x$  degree of famousness, and  $x$  is greater than the degree of famousness to which any real person exemplifies being-a-famous-detective

- (c'') Rodion Raskolnikov<sub>2</sub> is an archetypal bundle of human properties.
- (F<sub>3</sub>) (a<sub>1</sub>'') Rodion Raskolnikov<sub>1</sub> does not exist.
- (a<sub>2</sub>'') Rodion Raskolnikov<sub>2</sub> does not exist except as an abstract entity.
- (b<sub>1</sub>'') Sherlock Holmes<sub>1</sub> does not exist.
- (b<sub>2</sub>'') Sherlock Holmes<sub>2</sub> does not exist except as an abstract entity.

### Interpretation under Fictional Fictionalism

- (F<sub>1</sub>) (a''') Rodion Raskolnikov<sub>1</sub> is a murderer in  $w_{CP}$ .
- (b''') Sherlock Holmes<sub>1</sub> plays the violin in  $w_{SH}$ .
- (F<sub>2</sub>) (a'') According to the realist's paraphrase, Sherlock Holmes<sub>2</sub> encodes being-a-clever-detective to  $x$  degree; whereas Porfiry Petrovich<sub>2</sub> encodes being-a-clever-detective to  $y$  degree of cleverness, and  $x > y$ .
- (b'') According to the realist's paraphrase, Sherlock Holmes<sub>2</sub> encodes being-a-famous-detective to  $x$  degree of famousness, and  $x$  is greater than the degree of famousness to which any real person exemplifies being-a-famous-detective.
- (c'') According to the realist's paraphrase, Rodion Raskolnikov<sub>2</sub> is an archetypal bundle of human properties.
- (F<sub>3</sub>) (a''') Rodion Raskolnikov does not exist.
- (b''') Sherlock Holmes does not exist.

### Interpretation under Collingwoodian Mentalism

- (F<sub>1</sub>) (a''') Rodion Raskolnikov<sub>1</sub> is a murderer.
- (b''') Sherlock Holmes<sub>1</sub> plays the violin.
- (F<sub>2</sub>) (a''') Sherlock Holmes<sub>1</sub> is a more clever detective than Porfiry Petrovich<sub>1</sub>.
- (b''') Sherlock Holmes<sub>1</sub> is more famous than any real detective.
- (c''') Rodion Raskolnikov<sub>2</sub> is an archetypal character.
- (F<sub>3</sub>) (a''') Rodion Raskolnikov does not exist independent of human cognition.
- (b''') Sherlock Holmes does not exist independent of human cognition.

Our third and final conclusion is that there is now a path open to the Collingwoodian mentalist for undermining the abstract realist theory of fiction, should such a mentalist so choose. In adopting a purely anti-realist theory of fiction, the Collingwoodian mentalist is still capable of being faithful to the project of fictionalism. With this in mind, such a mentalist must first reduce all discourses previously thought to incur ontological commitments (modal logic, mathematics, identity, causality, etc.) to discourses which occur within Collingwoodian/Waltonian pretenses. Such a construal of these pretenses will curtail their ontological commitments to nothing more than an anti-realist ontological commitment to private human emotions and sensations. Once this

(admittedly enormous) task is complete, the Collingwoodian mentalist may then return to the abstract realist's theory of fiction and firmly accuse it of violating Occam's razor: when abstract entities are eliminated from all other discourses, the very existence of a functional anti-realist theory of fiction will render all realist theories excessive in their ontologies.

## BIBLIOGRAPHY

- Adams, Fred, Gary Fuller, Robert Stecker. "The Semantics of Fictional Names." *Pacific Philosophical Quarterly*, 78 (1997): 128-148.
- Brock, Stuart. "Fictionalism about Fictional Characters." *Noûs* 36 (2002): 1-21.
- Braun, David. "Empty Names, Fictional Names, Mythical Names." *Noûs*, 34 (2005): 596-631.
- Braun, David. "Empty Names." *Nous* 27 (1993): 449-69.
- Carter, W. R. "Do Creatures of Fiction Exist?" *Philosophical Studies*, 38 (1980): 205-215.
- Collingwood, R. G. *The Principles of Art*. London, Oxford, New York: Oxford University Press, 1938.
- Dostoevsky, Fyodor. *Crime and Punishment*. Translated by Richard Pevear and Larissa Volokhonsky. New York: Random House, 1992.
- Evans, Gareth. "Can There be Vague Objects?" *Analysis* 38 (1978): 208.
- Everett, Anthony. "Against Fictional Realism." *Journal of Philosophy*, 102 (2005): 624-49.
- Lewis, David. "Truth in Fiction." *American Philosophical Quarterly*, 15 (1978): 37-46.
- McDowell, John. *World and Mind*. Cambridge, Mass.: Harvard University Press, 1994.
- Plantinga, Alvin. *Essays in the Metaphysics of Modality*. Edited by Matthew Davidson. Oxford: Oxford University Press, 2003.
- Quine, W. V. *Word and Object*. Cambridge, Mass.: MIT Press, 1960.
- Sorensen, Roy. "Fictional Incompleteness as Vagueness." *Erkenntnis*, 34 (1991): 54-72.
- Salmon, Nathan. "Nonexistence." *Noûs*, 32, 3 (1998): 277-319.
- Stanley, Jason. "Hermeneutic Fictionalism." *Midwest Studies in Philosophy*, 25 (2001): 36-71
- Taylor, Kenneth. "Emptiness without Compromise." in *Empty Names, Fiction, and the Puzzles of Nonexistence*. Edited by A. Everett, T. Hofweber. Stanford: CSLI Press, 2000. Pp. 17-36.
- Thomasson, Amie. *Fiction and Metaphysics*. New York: Cambridge University Press, 1998.
- van Inwagen, Peter. "Creatures of Fiction." *American Philosophical Quarterly*, 14, 4 (1977): 299-308.



Walton, Kendall. *Mimesis as Make-Believe*. London, England: Harvard, 1990.

Wittgenstein, Ludwig. *Philosophical Investigations*. Translated by G. E. M Anscombe, 3<sup>rd</sup> ed. Oxford: Blackwell Publishing, 2001.

Wright, Crispin. *Truth and Objectivity*. Cambridge, Mass. & London: Harvard University Press, 1992.

Yablo, Steven. "Go Figure: A Path through Fictionalism." *Midwest Studies in Philosophy* 25 (2001): 72-102.

## VITA

Joseph John Dartez, Jr., son of Joseph Dartez and Joni (Currie) Dartez, was born in May, 1984, in Metairie, Louisiana. He graduated from Kolbe Academy in 2002 and attended the University of Dallas later that year. He spent the spring semester of 2004 in Rome, Italy, to study ancient Roman art, architecture and culture. During the summer of 2004, he attended a summer research program at The College of Wooster, where he modeled special cases of the Three-Body Problem. He married Sarah Hamilton, daughter of Edward Hamilton and Gaye (Galloway) Hamilton on January 8, 2005. In May, 2006, he graduated *cum laude*, earning a Bachelor of Arts in philosophy with a minor in applied physics, and in August, 2007, he entered the Graduate School at Louisiana State University.