Fiction and Fictionalism

Sofía Meléndez Gutiérrez

Queens' College, University of Cambridge

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Fiction and Fictionalism
Sofía Meléndez Gutiérrez

Abstract

A fictionalist theory is a theory that concerns a particular sphere of discourse, and is meant to encompass three theoretical virtues: a naturalistic ontology; an anti-revisionary attitude concerning our discursive practices; and the ability to foster semantic uniformity across discursive spheres. There are compelling reasons to believe that fictionalist theories are superior to their rivals; and, due to the promise that these theories appear to bear, I believe that it is worth inquiring further, and more systematically than it has been done in the past, whether or not they are truly viable. In this thesis, I will contend that they are not. The thesis comprises two parts. Part One contains an argument that calls into question the tenability of fictionalist theories in general. In order to advance this argument, I will discuss three of the extant theories of the ontology of *ficta*; I will advance a novel proposal thereof, and I will articulate an original solution to a prominent puzzle within the philosophy of music: the audibility problem. Part Two comprises discussions of three distinct fictionalist theories: mathematical fictionalism, modal fictionalism and moral fictionalism. I will argue that these three theories are all objectionable on the very same grounds, and I will thereby suggest that fictionalist theories in general are flawed in a systematic manner.

Acknowledgements

Si (como afirma el griego en el *Crátilo*)
el nombre es arquetipo de la cosa
en las letras de 'rosa' está la rosa
y todo el Nilo en la palabra 'Nilo'.

[If (as the Greek claims in the *Cratylus*)
the name is the archetype of the object
in the letters of 'rose' there is the rose
and the whole Nile in the word 'Nile'].

El Golem,
Jorge Luis Borges, 1964.

Contents

Introduction

Let us say that an *exotic entity* is a *non-naturalistic entity:* an entity that does not belong to the realm of the physical, and is such that its existence cannot be inferred nor verified by the methods of empirical science. Abstract entities, concretist and *ersatz possibilia*, moral properties and Meinongian objects, then, are all exotic.

In everyday life, we often seem to refer to exotic entities, and quantify over them. We do so, for instance, when we say that two is the only even prime, or that torture is abominable. If the possible world analysis of modal discourse is right, moreover, we do so also when we say things like the following: 'Diana Damrau's 2017 performance of Mozart's *Queen of the Night* could not have been more virtuosic'. And yet, the existence of exotic entities has been systematically and vehemently denied by many on the grounds that it contravenes the naturalistic conviction that all existing things exist within the reach of empirical science.

Fictionalism is a prominent method for vindicating this naturalistic conviction. Let K be a kind of exotic entities, and let K-sentences be the sentences that appear to involve reference to Ks or existential quantification over Ks. Suppose, additionally, that we assert K-sentences regularly in everyday life. A fictionalist theory about K-sentences is a theory that encompasses the following three doctrines:

Anti-Realism
Ks do not exist.

Standard Semantics
K-sentences are to be read at face value: as their surface form suggests, they involve purported reference to Ks, or quantification over Ks. Since Ks do not exist, K-sentences are untrue.

Anti-Revisionism

For every K-sentence s, s is typically not used to assert that which s literally expresses. Instead, s is ordinarily used non-literally to convey a piece of veridical information about the naturalistic world. Consequently, we may rationally use K-sentences as we always have—even if they are untrue.

Prima facie, there are compelling reasons to believe that fictionalism about K-sentences is superior to all rival theories. These rivals are three: realism, nominalism and eliminativism. Realism states that K-sentences refer to Ks and quantify over Ks. It states, moreover, that Ks exist, and they are exactly as described by K-sentences. Thus, on the realist's view, K-sentences are true, and it is rational to put them forward as we do.

The fictionalist contends that fictionalism is superior to realism because fictionalism provides us, at no cost, with a preferable ontology. Like realism, fictionalism construes K-sentences at face value, and states that it is rational to use them as we do—but realism comprises a commitment to a realm of exotic entities, whilst fictionalism does not. Thus, the ontology of the fictionalist is more modest; and, since exotic entities are non-naturalistic by definition, the fictionalist's ontology is the only one of the two that is naturalistic.

The core tenet of nominalism is that there are no Ks, but K-sentences are true nonetheless. In order to uphold this tenet, the nominalist attributes a non-standard semantics to K-sentences by way of one of two alternative theses. The first says that K-sentences are not to be read at face value: that, despite appearances, they do not involve reference to Ks, nor quantification over Ks. The second, by contrast, states that K-sentences are to be read at face value—but it also states that they possess non-standard truth conditions, for they are true despite the fact that Ks do not exist. Since K-sentences are true, the nominalist concludes, it is rational to use them as we do.

The fictionalist maintains that her theory surpasses nominalism; she maintains, more specifically, that fictionalism delivers all the same goods as nominalism—but does so in a theoretically more virtuous manner. Both nominalism and fictionalism are free of commitments to Ks, and both construe our use of K-sentences as rational. Nevertheless, fictionalism attributes a standard semantics to K-sentences, whilst nominalism does not; and, consequently, fictionalism, but not nominalism, fosters semantic uniformity across the various discursive spheres that exist.

Eliminativism is the sum of (Anti-Realism), (Standard Semantics), and the following doctrine:

Revisionism

Since asserting untrue statements is irrational, asserting K-sentences is irrational. Rationality thus compels us to abandon our use of K-sentences.

The fictionalist claims that her theory is preferable to eliminativism for one simple reason: it allows us to preserve our use of K-sentences without irrationality and at no cost—for, within the K-fictionalist's framework, we may rationally assert K-sentences without committing to Ks, and without attributing a non-standard semantics to these sentences.

In sum, then: the project of the K-fictionalist is the project of encompassing three theoretical virtues: a naturalistic ontology; an anti-revisionist policy concerning our discursive practices; and the ability to foster semantic uniformity across discursive spheres. As a consequence of this, fictionalism seems, *prima facie* at least, to surpass all of its rivals.

Despite their notorious attractiveness, objections have been raised against virtually all of the fictionalist theories that exist. Fictionalists have generally contested these objections with plausibility; and yet, given the remarkable promise that these theories appear to bear, I believe that it is worth inquiring further, and more systematically than it has been done in the past, whether or not they are truly viable. In this thesis, I will contend that they are not. My thesis comprises two parts. The first contains an argument that calls into question the tenability of fictionalist theories in general. The second comprises three disjoint discussions of three particular theories.

Part One

Let us say that a *fictive sentence* is a sentence that appears to involve reference to *ficta*, or quantification over *ficta*.[1] The following, then, is a fictive sentence:

1.   Ireneo Funes is incapable of oblivion.[2]

———————————————

1. Throughout this thesis, I will use the term '*fictum*' as synonymous with 'fictional entity', and '*ficta*' as the plural of '*fictum*'.

2. Most of the fictive sentences that I use as examples in this thesis concern *Funes, el Memorioso* (Borges,

One of the fictionalist theories that exist concerns fictive sentences. Let us call it 'fictive fictionalism'. To my mind, fictive fictionalism is especially interesting amongst extant fictionalist theories. Let us see.

Let K be a kind of exotic entities, and let K-sentences be the sentences that appear to involve reference to Ks or existential quantification over Ks. Finally, let 'K_1' purport to refer to a particular K, and consider the following K-sentence:

2. K_1 is G.

By (Anti-Realism), K_1 does not exist. By (Standard Semantics), (2) is to be read at face-value: as its surface form suggests, it involves purported reference to K_1. Since K_1 does not exist, (2) is untrue. By (Anti-Revisionism), however, (2) is ordinarily used to convey truthful information. If this is the case, the fictionalist notes, then (2) is exactly alike (1): (1), the fictionalist claims, involves purported reference to an entity that does not exist, and it is, therefore, not true. And yet, we ordinarily use it to convey truthful information, for, ordinarily, we use it to express the following:

3.   According to Borges's *Funes, the Memorious*, Ireneo Funes is incapable of
         oblivion.

By the K-fictionalist's own lights, then, K-sentences are importantly analogous to fictive statements. They are so much so, in fact, that the K-fictionalist suggests that K-sentences are a species of fictive statements (see, e.g., Field, 1989, pp. 2–3; Rosen, 1990, pp. 230–232; Kalderon, 2005, pp. vii–x).

Unfortunately, the suggestion in question delivers a catastrophic result. If K-sentences

_____

1944a) (i.e., *Funes, the Memorious*), a short story written by Jorge Luis Borges in 1942, and published a his 1944 collection *Ficciones* (i.e., *Fictions*). For the reader who is unfamiliar with this story, here is a summary: Ireneo Funes is a nineteen-year-old Uruguayan youth who became insomniac and paralysed after an equestrian accident. After the accident, however, he also acquired extraordinary perceptual abilities and a prodigious memory: he perceives his surroundings in overwhelming detail, and he is radically incapable of forgetting anything of what he perceives. Borges's story narrates a conversation that Ireneo had with a man who lent him a couple of Latin texts and dictionaries. (Ireneo would then learn Latin from these texts in a matter of days.) During this conversation, Ireneo explains what it is to live with his unrelenting perception and memory: he says, for instance, that just as we can all recall in detail in our minds the shape of a perfect circle, he can thoroughly recall the shape had by a colt's tumultuous mane at any given instant. He also explains that his world is inhabited by nothing other than extremely finely individuated particular entities: his relentless memory and perception preclude him from omitting details and, hence, they preclude him from abstracting a universal from its instances, and from identifying the various temporal and spatial stages of an individual.

are fictive statements, then, by the definition of 'fictive statement', the entities that K-sentences concern are fictional entities. The entities that K-sentences concern are, of course, Ks. Thus it follows that, on the K-fictionalist's view, Ks are fictional entities. This result is severely pernicious, however, for a wealth of authors have argued compellingly that our use of fictive sentences can only be honoured under the assumption that fictional entities are not non-existent, but exotic: creationists (e.g., Van Inwagen, 1977; 1983; 2000; Salmon, 1998; Thomasson, 1998; Voltolini, 2006; Kripke 2013) and Platonists (e.g., Wolterstorff, 1980, pp. 134–197; Friedell, 2020b, p. 224), for instance, believe that *ficta* are abstract; possibilists (e.g., Lewis, 1978; Stone, 2010; Caplan, 2016; Haraldsen, 2017; 2020) maintain that *ficta* are concretist *possibilia*; and noneists (e.g., Parsons 1980; Zalta, 1983, pp. 41–50: Jacquette, 1996) hold that they are Meinongian objects. If any one of these authors is right, then the K-fictionalist's claim that Ks are fictional entities is nothing other than the declaration of commitment to the existence of exotic entities of one sort of another. If any one of the authors in question is right, therefore, then K-fictionalism is straightforwardly inconsistent with the rejection of exotic entities (Brock, 2002, pp. 10–12; Eklund, 2019; Falguera, Martínez-Vidal & Rosen, 2022, §2.1).

I said earlier that the project of the K-fictionalist is that of encompassing three theoretical virtues: a naturalistic ontology; an anti-revisionist policy concerning our discursive practices; and the ability to foster semantic uniformity across discursive spheres. If K-fictionalism is indeed inconsistent with the rejection of exotic entities, then the project in question has failed: the K-fictionalist's ontology is not naturalistic after all. I will henceforth refer to this result as 'the failure result'.

Fictionalists have tried to block the failure result by advancing a fictionalist theory of fictive statements (see, e.g., Walton, 1990; Brock, 2002; 2016; Everett, 2005, pp. 638-649; 2013, pp. 6–103). The defenders of this theory contend that our use of fictive statements may be honoured *without* admitting fictional entities into our ontology. On their view, the K-fictionalist's claim that Ks are *ficta* is not at all a declaration of commitment to exotic entities of any sort: it is, instead, nothing other than a tortuous way of saying that Ks do not exist.

Now, I said earlier that fictive fictionalism is especially interesting amongst the fictionalist theories that exist. The reason why it is so is this: fictive fictionalism is the tool that fictionalists have employed to try to block the failure result.

In virtue of the especial interest that fictive fictionalism bears, Part One of my thesis will be devoted to this one theory. And more specifically: the six chapters that comprise Part One constitute an argument to the effect that fictive fictionalism cannot be rationally endorsed.

Everett's pretence theory is the most prominent amongst the various fictionalist accounts that exist of fictive statements. In Chapter One, I will advance two novel objections and a dilemma against it. I will thereby conclude that the project of accounting for fictive statements within a fictionalist framework is not promising. Accordingly, in Chapter Two and Chapter Three, I will consider the fictive fictionalist's rivals: I will consider, that is to say, two mutually competing, realist accounts of fictive statements: creationism and possibilism. I will show, however, that these two theories, too, are unsuccessful. Consequently, I will advance, at that point, an original proposal: the type theory. The type theory is a realist theory: it states that fictive statements are true in virtue of the existence of fictional entities. On the type theorist's view, moreover, *ficta* are types of concretist *possibilia*—types, that is to say, of concretist non-actual individuals. Throughout Chapter Four, Chapter Five and Chapter Six, I will formulate, defend and motivate this view: I will show that it is parsimonious and explicatively powerful, and I will also demonstrate that it is uniquely fit for the project of assigning all sorts of non-concrete artistic creations (i.e., *ficta,* musical works, literary works, drama, etc.) to one and the same ontological category. Finally, in the process of articulating the demonstration in question, I will propose an original solution to a prominent puzzle within the philosophy of music: the audibility problem.

If my arguments in Part One are sound, then there is a realist theory of fictive statements that determinately surpasses fictive fictionalism. Fictive fictionalism, therefore, cannot be rationally endorsed. For fictionalists, this result is seriously detrimental: as I explained earlier, fictive fictionalism is the tool that fictionalists have employed to try to block the failure result. If fictive fictionalism cannot be rationally endorsed, then there is currently no tool for fictionalist to counter said result. Surely other tools could be developed—but that would not suffice to neutralise my argument.

Fictive discourse is, as a matter of fact, a discourse that we recurrently engage with in everyday life. If my arguments in Part One are sound, then our engagement with it cannot be accounted for within a fictionalist framework: in order to honour said engagement, we ought to postulate *ficta*, and identify them with types of concretist

possibilia. Types of concretist *possibilia* are, so to speak, doubly exotic: they themselves are abstract, but their identity conditions ought to be formulated in terms of concretist *possibilia.* If my arguments are sound, therefore, we are, by our use of fictive statements, committed to not one, but two sorts of exotic entities.

Two interesting results follow, then, from my type theoretic account of fictive discourse. The first is that mathematical fictionalism and modal fictionalism are futile. As I will explain in Chapter Seven and Chapter Eight, mathematical and modal fictionalists wish to account for mathematical and modal discourses in fictionalist terms in order to deliver us from committing, respectively, to *abstracta* and concretist *possibilia.* If *ficta* are types of concretist *possibilia,* however, then we are inescapably committed to both concretist *possibilia* and *abstracta* merely in virtue of our use of fictive statements. The second result concerns not mathematical and modal fictionalists only, but fictionalists in general. If in virtue of our use of fictive statements we are committed to two sorts of entities that are non-naturalistic, then, to my mind, the question arises of whether it is worth even trying to escape other sorts of non-naturalistic commitments. Is it worth, for instance, trying to escape a commitment to moral properties on the grounds that they are non-naturalistic if we will have non-naturalistic commitments anyway? Or might it be best to call into question our naturalistic convictions?

Part Two

Part Two is constituted by Chapter Seven, Chapter Eight and Chapter Nine, and it comprises discussions of three distinct fictionalist theories: mathematical fictionalism, modal fictionalism, and moral fictionalism. Throughout the chapters that comprise Part Two, I will advance a novel objection against every one of these theories. Importantly, moreover, the core contention behind each one of these objections will be one and the same. More specifically: for each one of these objections, if the objection concerns K-fictionalism, then the objection states that the K-fictionalist cannot uphold an anti-revisionist policy with regard to K-sentences—not, at least, without definitively renouncing anti-realism about Ks, or a standard semantic account of K-sentences. In any case, I will conclude, the K-fictionalist project cannot be carried through. Importantly, moreover, the fact that the three theories in question can be objected to on the same grounds suggests that perhaps fictionalist theories in general are flawed in a systematic manner.

14

Part One

Chapter One:
Fictive Fictionalism

Fictive sentences are those that involve purported reference to *ficta*, or quantification over *ficta*. Witness:

1.   Ireneo is a hypermnesiac.
2.   Ireneo was created by Jorge Luis Borges in 1942.
3.   Some fictional entities are more realistic than others.

Some fictive sentences are intuitively true, and these fictive sentences seem to give raise to an unpalatable dilemma. Consider, for instance, (1)–(3). (1)–(3) are intuitively true. If they are true indeed, then the name 'Ireneo' refers, and the existential quantifier in (3) ranges over a domain that comprises fictional entities. If (1)–(3) are true, therefore, then we ought to conclude that there are fictional entities—but this conclusion seems unacceptable: it appears platitudinous that, for every name 'a', the statement that a is fictional sanctions the conclusion that a does not exist. On the other hand, if fictional entities do not exist, then we ought to conclude that (1)–(3) are untrue. If they are untrue, then we are rationally compelled to revise our use of them: on pain of irrationality, (1)–(3) must never be asserted again. This result is undesirable, however: asserting fictive sentences like (1)–(3) is crucial for us to describe and discuss the contents of works of literature, and to do literary criticism.

It appears, in sum, that intuitively true fictive sentences produce an unpalatable dilemma: they force us to choose between an undesirable sort of realism and an undesirable sort of revisionism. Fictive fictionalism (which I will call simply 'fictionalism' in this chapter) is a fictionalist theory fictive statements, and it states that the dilemma in question need not be faced. On the fictionalist's view, we may rationally preserve our use of fictive sentences without committing to the existence of fictional entities. In other words: the

fictionalist project consists in encompassing an anti-realist theory of fictional entities with an anti-revisionary policy concerning our use of fictive statements.

Everett's pretence theory is a prominent instance of fictionalism: it is significantly more elaborate than it predecessors (e.g., Evans, 1982; Walton, 1990; 2000; Crimmins, 1998; Kroon, 2000; 2004), but it draws judiciously on them. In this chapter, I will advance two original objections and a dilemma against this theory. I hope to thereby establish that the prospects for success of the fictionalist project are poor.[3]

In §1, I will introduce Everett's theory, and present a notable complaint that has been raised against it: the phenomenological objection. In §2, I will reconstruct Everett's response to this objection, and, in §3, I will argue against this response by way of two independent arguments. Finally, in §4, I will formulate a dilemma for Everett. One of its horns leads him to the conclusion that his theory fails the fictionalist purpose of fostering anti-revisionism; the other one leads him to a daunting challenge. I will thus conclude that Everett's theory is not satisfactory.

1   Pretence and the Phenomenological Objection

Earlier I described a dilemma whose horns are an undesirable form of realism, and an undesirable form of revisionism. This dilemma, I said, seems to arise in connection with intuitively true fictive statements—e.g., (1)–(3) The fictionalist theory that constitutes the subject matter of this chapter is an attempt to dissolve this dilemma and, accordingly, I will be concerned, in what follows, not with fictive statements in general, but with those fictive statements that appear to us to be true. Thus, for the sake of simplicity, I will henceforth use the term 'fictive sentences' to refer exclusively to intuitively true fictive statements.

The monograph *The Nonexistent* (2013, pp. 6–103) comprises Everett's most elaborate rendition of his fictionalist theory, and a very illuminating outline was published earlier, in his (2005, pp. 638–649). I will refer to both of these works in what follows.

According to Everett, fictive sentences are all untrue because fictional entities do not exist. We put these sentences forward nonetheless—but never sincerely: the attitude that relates us to fictive sentences is not belief, but make-belief, and, accordingly, when

---

3. Brock's fictionalist theory (2002; 2016) is prominent, too, but it has been thoroughly criticised by Zouhar (2018), and, consequently, I shall not discuss it.

we assert fictive sentences we do so only pretencefully. On Everett's view, then, we do not err when we advance fictive sentences, just as Laurence Olivier does not err when he, on stage, exclaims: 'This is I, Hamlet the Dane'. Importantly, moreover, make-belief is ontologically non-committing: one may hold a stick and make-believe that it is a sword without committing to the existence of a sword that one is holding. Analogously, we may make-believe that Borges created Ireneo in 1942 without committing to Ireneo's existence. In sum, therefore, Everett contends that we may utter fictive sentences as we always have without error, and without admitting fictional entities into our ontology.[4]

An objection suggests itself immediately: phenomenologically, an assertion of a fictive sentence is not at all akin to a pretenceful assertion of an untrue statement. According to the understanding that we possess of our own attitudes, when we assert a fictive sentence, we do so sincerely and, accordingly, the attitude that relates us to fictive sentences is nothing short of genuine belief. Unless we err systematically in interpreting our own attitudes, therefore, Everett's theory is false.[5]

Everett formulated several rejoinders to this so-called 'phenomenological objection'. The first says that our understanding of our own mental phenomena is, in general, rather poor: false memories, for instance, are indistinguishable from veridical ones, and no amount of introspection suffices to identify them as false. Introspection, then, is unreliable as a guide for understanding our own mental phenomena, and, consequently, our convictions that the attitude that relates us to fictive sentences is genuine belief should not be taken too seriously (Everett, 2013, pp. 110–112). Moreover, Everett (2013, pp. 116–119) maintains, the pretence theory of fictive sentences is the result of a faultless abduction: we are thoroughly convinced that fictional entities do not exist, and yet we speak as if they did—consider, for instance, (1)–(3). The best way to make sense of this datum, Everett insists, consists in supposing that the relevant pieces of discourse are insincere. Everett's principal response to the phenomenological objection, however, is a sophisticated explanation of why we mistake our utterances of fictive sentences for sincere assertions of true statements. I will present this explanation in the following

---

4. One may wonder what it is to make-believe that Borges created Ireneo if Ireneo does not exist and, consequently, the name 'Ireneo' does not refer. Everett maintains that we can entertain thoughts that lack a referent; that we can, for instance, think about Ireneo even if the name 'Ireneo' does not refer (2013, p. 4). On this view, we can make-believe that Borges created Ireneo even if the name 'Ireneo' lacks a referent. The question whether this view is correct falls outside the scope of this chapter.

5. Everett attributes this objection to Stanley (2001), and discusses it, along with other objections, in his (2013, pp. 103–119).

section.

2   Base Pretences and Extended Pretences

The fictive sentences that appear to us to be true, like (1)–(3), come in various kinds. Some of them are *internal* or, more precisely, *internal relative to a certain work of fiction:* for any fictive sentence s, s is internal relative to a work of fiction F just in case s is true according to F. Consider, for instance, (1): (1) is true according to Borges's *Funes, the Memorious*, so (1) is internal relative to *Funes, the Memorious*.

According to Everett (2005, p. 638; 2013, pp. 30–37), engaging imaginatively with a work of fiction F requires us to enter a pretence P_F where all sentences that are internal relative to F are true. Everything that is true in F, then, is true in P_F: all of the persons, creatures, objects, places and events that exist in F exist in P_F, and they are exactly as F describes them. On this view, engaging imaginatively with *Funes, the Memorious*, for instance, requires us to enter a pretence where Ireneo exists, he is a hypermnesiac, he lives in Fray Bentos, in Uruguay, he had an equestrian accident, etc. Pretences like this one Everett calls 'base pretences' (2013 p. 34). Hereafter, let us say that *the base pretence associated with a work of fiction F* is the base pretence where all the sentences that are internal relative to F are true.

Describing or discussing the contents of a story often requires us to assert internal sentences. We may need to assert (1), for instance, when describing or discussing the contents of *Funes, the Memorious*. Everett (2005, p. 639; 2013, pp. 46–53) claims that asserting a sentence that is internal relative to a work of fiction F requires us to enter the base pretence associated with F. Consequently, on his view, internal sentences are always asserted within pretences; they are always asserted pretencefully.

A fictive sentence s, we have seen, is internal relative to a work of fiction F just in case s is true according to F. Hence, if s is internal relative to a F, and s describes a fictional entity that figures in F, then s describes that fictional entity in accordance with F. Not all of the fictive sentences, though, describe *ficta* in accordance with the stories where they figure: some of them, instead, describe *ficta* from the perspective of the literary critic who sees fictional entities as authorial creations, and judges them aesthetically as such. These *critical* sentences say, for instance, that *ficta* were brought into existence by writers, and that they possess aesthetic qualities such as being realistic, well-constructed, original, revolutionary, and typical or exemplar of a certain literary

tradition. Thus (2) and (3) are critical sentences.[6]

Everett suggests that critical sentences constitute a peculiar fiction created collectively by literary critics; a fiction that states that there are fictional entities, and that they are the products of the creative activities of writers, and objects of aesthetic appreciation. Moreover, Everett believes that our assertions of critical sentences are always pretenceful: they pertain, Everett maintains, to a natural extension of the pretences that we enter when we imaginatively engage with works of fiction (2005, pp. 639–640; 2013, pp. 53–63). Consider:

4.   Ireneo is a hypermnesiac man that Borges created in 1942.

In order to assert (4), Everett claims, we enter the base pretence associated with *Funes, the Memorious*: we make-believe that Ireneo exists and that he is a hypermnesiac—and then we also make-believe that he was brought into existence by Borges in 1942. Thus, in asserting (4), we extend our base pretence in order to incorporate the critics' fiction (Everett, 2005, pp. 639–640; 2013, pp. 53–63). The result is an *extended pretence* (Everett, 2013, p. 54).[7] Once our pretence has been extended, we may, by uttering critical sentences, speak freely of fictional entities as if they were authorial creations and objects of aesthetic appreciation.

Following Evans (1982) and Walton (1990), Everett (2005, p. 641; 2013, p. 18) claims that every pretence is governed by *principles of generation:* principles that dictate what is true in the pretence. Some of these principles are *bridging principles*—principles that state that a sentence s_1 is true in the pretence just in case a sentence s_2 is true in reality (Everett, 2005, p. 641; 2013, pp. 18–26). Consider, for instance, the fencing game, which is governed by the following bridging principle:

Fencing
'x caused a deadly injury to y' is true in the pretence $\leftrightarrow$ 'x touched y' is true in reality.

Everett thinks that we have an intuitive understanding of the bridging principles that

---

6. There are fictive sentences that are neither internal nor critical, but I shall not discuss them in this chapter. See, e.g., Sainsbury (2009, pp. 122–126) and Button (2012) for interesting discussions of fictive sentences of other kinds.

7. In his 2005, Everett uses other terms to designate base pretences and extended pretences: he calls them 'P_1 pretences' and 'P_2 pretences' respectively.

govern the pretences that we participate in. As a consequence, he maintains, pretenceful assertions convey truthful information about the real world at the level of pragmatics (2005, pp. 41–46; 2013, pp. 38–46). Suppose that a is participating in the fencing game, and touches b. She triumphantly exclaims that she has caused a deadly injury to b. We all understand that the fencing game is governed by (Fencing) and, consequently, a's exclamation pragmatically conveys the truthful piece of information that she has touched b in the real world. The mechanism whereby pretenceful assertions convey information Everett calls '*piggybacking*' (2013, p. 38). In schematic terms, then, 'piggybacking' is to be defined as follows:

> **Piggybacking**
> Let P_F be a pretence. Then:
>> *i.* An assertion of a sentence s_1 within P_F piggybacks truthful information $\leftrightarrow$ there is a sentence s_2 such that the assertion of s_1 pragmatically conveys the information that s_2 is really true.
>> *ii.* (A pretenceful assertion of s_1 within P_F pragmatically conveys the piece of truthful information that s_2 is really true) $\leftrightarrow$ (P_F is governed by a bridging principle stating that: s_1 is true in P_F $\leftrightarrow$ s_2 is true in reality).

Everett says that base pretences, and extended ones, too, are governed by principles of generation. Consequently, assertions made within pretences of both sorts convey truthful information about the real world. Base pretences are governed by one bridging principle that is easy to conjecture. Earlier I said that a base pretence is that in which we make-believe everything that is true in a given story. Hence, if P_F is the base pretence associated with a work of fiction F, then the principle governing P_F is the following:

> **Base Pretence**
> s is true in P_F $\leftrightarrow$ in the real world, s is true according to F.

If this is the case, then, by (Piggybacking), an assertion of s in P_F conveys the piece of truthful information that, in the real world, s is true according to F (Everett, 2013, pp. 50–52). Consider (1), for instance. There is a base pretence associated with *Funes, the Memorious*. By (Base Pretence), (1) is true in this pretence iff, in the real world, (1) is true according to *Funes, the Memorious*. By (Piggybacking), an assertion of (1) within the pretence conveys truthful information: it conveys that, in the real world, (1) is true according to *Funes, the Memorious*. A pretenceful assertion of (1), therefore,

piggybacks the following piece of truthful information:

5.   According to *Funes, the Memorious*, Ireneo is a hypermnesiac.[8]

On Everett's view, extended pretences, too, are governed by bridging principles that allow us to piggyback truthful information. Let us consider an example: earlier I noted that (2) is a critical sentence, and critical sentences, on Everett's view, are asserted within extended pretences. Given the bridging principles that govern extended pretences, a pretenceful assertion of (2) conveys, roughly, the following piece of truthful information (Everett, 2013, pp. 58–60):

6.   Borges created a story in 1942, and, according to this story, Ireneo exists, but 'Ireneo exists' is not true in reality.

On Everett's view, in sum, our assertions of fictive sentences, internal and critical alike, occur within pretences. Nevertheless, Everett thinks that the pretences where these assertions take place differ from other pretences in an important respect. An utterance made within an ordinary make-believe game is *game-oriented:* it constitutes a move that the speaker makes for the sake of bringing about a development of the game itself. In the fencing game, for instance, *a*'s exclamation that she caused a deadly injury to *b* is meant to establish that, within the make-believe world where *a* and the other participants are fencers, *b* counts, from then on, as defeated. By contrast, utterances of fictive sentences are *world-oriented:* they are made in order to exploit their piggybacking properties or, equivalently, in order to convey truthful information about the real world. According to Everett, it is hard, or even impossible, to convey the particular pieces of

---

8. Intuitively, it is unquestionable that (5) is true, but, in fact, it is rather unclear how it can be true within the framework of Everett's theory. Let us say that any sentential operator of the form 'according to *F*', where *F* is a work of fiction, is a *fiction operator*. Let us say, moreover, that a statement governed by a fiction operator is a *parafictive statement*. Fiction operators are clearly circumstance-shifting: a sentence that is untrue may be true in a work of fiction, so a sentence of the form 'according to *F*, *s*' may be true even if *s* is untrue. Everett, however, maintains that fictive and non-fictive names are devices for direct reference: the semantic value of a name is exhausted by its referent, and does not include a Fregean sense or a descriptive content of any sort (2013, p. 3). If this is the case, and *ficta* do not exist, then the name 'Ireneo' in (1) is wholly semantically empty. (1), then, comprises a semantic gap, and this semantic gap cannot be filled by embedding (1) within the scope of the fiction operator: the fiction operator may be circumstance-shifting, but it simply does not have the powers to fill semantic gaps. If this is right, then (5) (the parafictive statement that results from embedding (1) within the scope of the fiction operator) comprises a semantic gap, too. The question then arises of how (5) is true. An interesting answer to this question may be found in Friend's work (2011; 2014), but it falls outside the scope of this chapter to determine whether or not Friend's answer is sound. In what follows, I shall simply assume, for the sake of Everett's argument, that (5) is true even if 'Ireneo' is empty—and, more generally, that a parafictive statement may be true even if it comprises an empty fictive name.

truthful information that are conveyed by pretenceful assertions of fictive sentences by means other than piggybacking: in order to convey these pieces of information, that is to say, one ought to enter the appropriate pretence, assert the appropriate fictive sentence, and let pragmatics do its trick. Our pretenceful assertions of fictive sentences, therefore, serve the very serious purpose of conveying truthful information that cannot be conveyed by other means—and that, Everett contends, is why we mistake the assertions in question for genuine assertions of true statements (Everett, 2005, pp. 645–648; 2013, pp. 2, 54, 117).

3   Two Objections

Let us grant Everett, for now, that the piggybacking thesis explains away our introspection-grounded impressions that our assertions of fictive sentences are sincere. Sadly, there is linguistic evidence independent of these impressions to the effect that Everett's pretence theory is wrong. Unless this evidence is discredited, I submit, Everett's theory should be rejected. Consider:

7.   Ireneo suffers from insomnia.

(7) is true according to *Funes, the Memorious*, so it is internal relative to *Funes, the Memorious*. Everett, we have seen, believes that an assertion of an internal sentence *s* is made within the base pretence that is associated with the work of fiction that *s* is internal to. Consequently, assertions of (7) are made within the base pretence associated with *Funes, the Memorious*. On Everett's view, then, in order to assert (7), we ought to enter the pretence where everything that is narrated in *Funes, the Memorious* is true: we ought to make-believe that Ireneo exists, and that he is exactly as depicted in *Funes, the Memorious*—he is a hypermnesiac, he suffers from insomnia, etc. Now, if asserting (7) requires us to enter this pretence, then, surely, after asserting (7), we may felicitously keep speaking as if Ireneo existed. This prediction fails, however. Witness:

8.   Ireneo suffers from insomnia, and I hope he will soon manage it: it is known to cause memory impairments.

An utterance of (8) does not seem felicitous at all: its first half is intuitively true, but whoever asserts the second appears to be under the serious misunderstanding that Ireneo exists in the real world. Clearly, this suggests that the reason why (8) is infelicitous is that something goes awry right at the point where the speaker starts uttering the

second half of (8): she speaks as if Ireneo existed, even though he does not. If this is correct, then the speaker was not pretending that Ireneo exists when she asserted the first half of (8). Since the first half of (8) is (7), it follows that Everett is wrong: we need not pretend that Ireneo exists in order to assert (7).

In order to block this result, Everett could contend that the reason why an utterance of (8) is infelicitous is not the one that I have proposed, but the fact that the speaker who utters (8) speaks of Ireneo as if he were presently alive. A speaker who speaks like this, Everett could claim, sounds odd because Borges's story locates Ireneo's birthdate almost two centuries ago, in 1868.

Unfortunately, the aforestated claim cannot be right. Everett maintains that, in order to assert (7), we must enter a pretence where everything that is the case in *Funes, the Memorious* is the case. In *Funes, the Memorious*, the current year is 1887, and, according to *Funes, the Memorious*, Ireneo was alive in 1887. On Everett's own view, therefore, asserting (7) requires us to enter a pretence where Ireneo is alive at present. By his own lights, then, a speaker who asserts (7) speaks as if Ireneo were presently alive—and yet, (7) is not infelicitous. It follows that speaking of Ireneo as if he were alive presently is not infelicitous, and, accordingly, Everett cannot explain the infelicity of (8) by noting that the speaker who utters (8) speaks as if Ireneo were alive at present.[9]

Perhaps, though, what is off about (8) is that its first half is world-oriented, whilst its second half is game-oriented. Perhaps, that is to say, the first half of (8) serves the purpose of conveying worldly information by way of a piggybacking process, but its second half does not: its second half is a mere move within the pretence that Ireneo exists—a mere move within a make-believe game, just like *a*'s exclamation, in the fencing game, that she wounded *b* fatally. Thus Everett could contend that the second half of (8) sounds odd because it is a wholly ludic utterance that succeeds a serious, truthful and informative assertion. This contention, however, is untenable. If the second half of (8) is made within a pretence, then it is world-oriented: it is one of those pretenceful assertions that are made in order to convey truthful information for, evidently, it is

9. Some authors (e.g., Rosen, 1990, pp. 230–232; Salmon, 1998, p. 303; Kripke, 2013, pp. 57–58) would claim that the speaker who utters (7) does not speak as if Ireneo were presently alive. On their view, (7) is ordinarily used as elliptical for the following statement: 'according to *Funes, the Memorious*, Ireneo is a hypermnesiac'. If (7) is elliptical for this statement, then the speaker who utters (7) does not speak as if Ireneo were alive presently: she merely describes the content of Borges's story. Nevertheless, Everett (2013, pp. 46–53) explicitly rejects the ellipsis account of (7). On his view, therefore, it is definitively the case that the speaker who utters (7) speaks as if Ireneo were presently alive.

meant to convey information about a real emotional attitude that the speaker possesses: she hopes, that is to say, that Ireneo will manage his insomnia.[10] Hence, the orientation of both of the halves of (8) is the same: they are both world-oriented and, consequently, there is no reason why the orientation of the second one should cause it to sound odd when uttered right after the first.

In sum: if my arguments are sound, then Everett's account of internal sentences is severely dubious: it is severely dubious, that is to say, that our assertions of these sentences are pretenceful as Everett claims that they are.

Importantly, I believe that Everett's account of critical sentences, too, is defective. We have seen that Everett maintains that asserting a critical sentence requires entering a base pretence, and extending it so that it incorporates the fiction that fictional entities are authorial creations and objects of aesthetic appreciation. This, Everett (2005, p. 640) notes, implies that, in extended pretences, fictional entities are ontologically queer. Suppose, for instance, that we enter the base pretence associated with *Funes, the Memorious.* We make-believe that Ireneo exists, and that he is just as the story portrays him: he is Uruguayan, he is a tobacco smoker, he is the child of María Clementina Funes and an unknown father, etc. Then we extend our pretence by incorporating the critics' fiction. Thus we make-believe that Ireneo came into existence in 1942 as a product of Borges's literary endeavours. In our extended pretence, then, Ireneo is both a flesh-and-blood Uruguayan man and an object of aesthetic appreciation; he is the offspring of two humans, and Borges's authorial creation. In Everett's own words: he is "dual-natured" (2005, p. 640).

Everett believes that the thesis is innocuous that fictional entities are ontologically queer in extended pretences. The advocates of the pretence theory are not committed to the existence of dual-natured entities, but merely to the truth of the thesis that, under certain circumstances, we pretend that dual-natured entities exist. Pretending that entities of a certain class exist, Everett maintains, is innocuous regardless of how queer such entities are imagined to be (2013, p. 213). In what follows, however, I will argue that Everett is wrong—that the dual nature thesis is not innocuous, but severely

---

10. It might appear that a subject cannot really have an emotional attitude concerning Ireneo if she does not believe that Ireneo exist—but this is not the case. The phenomenon whereby persons possess emotional attitudes concerning entities that they know to be fictional is well recognised, and it has been extensively philosophically discussed. For an outline of the relevant discussions, see Liao & Gendler (2020).

detrimental for Everett's account of critical sentences.

Consider (2) again: (2) is a critical sentence. Everett states that, in order to assert (2), we ought to enter the base pretence associated with *Funes, the Memorious*, and extend it so that it incorporates the critics' fiction. The extended pretence where assertions of (2) occur, then, is a pretence where Ireneo is dual-natured: he is Borges's authorial creation, but he is also a flesh-and-blood Uruguayan man. If this is the case, then, in the pretence in question, the entity that was created by Borges according to (2) is a flesh-and-blood man. It follows that, in this pretence, Borges is capable of creating human beings. This conclusion is supported by Walton, on whose work Everett grounded his own theory: according to Walton (1990, p. 410), when we say that a writer created a fictional entity, we enter a pretence where we make-believe that, when a writer writes a fiction about a person, she brings said person into existence. To my mind, however, Walton's contention involves a deep misunderstanding of our assertions. When we say that a writer created a fictional entity, we acknowledge this writer's ability to produce literary objects, and do not attribute to her (not even pretencefully) any extraordinary abilities to create people. This is confirmed by the felicity of utterances of sentences like the following:

9.   Borges is a brilliant writer: he created Ireneo, one of the most memorable characters in Latin American literature.

If this is right, then it is severely implausible that, whenever we say that a *fictum* was created by a writer, we enter a pretence where this *fictum* is dual-natured: not only a literary object, but also a flesh-and-blood human—or whatever it is in the story where it figures. Since fictional entities are dual-natured in extended pretences, it is implausible that we enter extended pretences when we assert that a *fictum* was created by an author—but that is precisely what Everett maintains: he maintains that critical sentences, including those that concern the creation of fictional entities, are always uttered within extended pretences. Thus I conclude that the dual-nature thesis is, *pace* Everett, pernicious for his account of critical sentences: it forces him to blatantly misconstrue our assertions.

I submit, in sum, that, until both of the objections that I advanced in this section are successfully countered, Everett's theory cannot be considered satisfactory.

26

4   The Dilemma

Earlier I defined 'piggybacking' by way of (Piggybacking). Nevertheless, if this is how 'piggybacking' is defined, then Everett's contention that our pretenceful utterances of fictive sentences piggyback worldly information is inconsistent with one crucial tenet of his theory. Everett, we have seen, maintains that we mistake our pretenceful utterances of fictive sentences for sincere assertions of truthful statements—and we do so because said utterances serve the purpose of conveying worldly information which may be difficult, or even impossible, to convey otherwise (2013, p. 54). This, however, cannot be true if our pretenceful utterances of fictive sentences piggyback worldly information.

Suppose that s_1 is a fictive sentence, and that, ordinarily, utterances of it occur within a pretence P_F. According to (Piggybacking), an assertion of s_1 within P_F piggybacks truthful information just in case there is a really true sentence s_2 such that the assertion in question pragmatically conveys that s_2 is really true. Suppose that a pretenceful assertion of s_1 within P_F piggybacks truthful information. By (Piggybacking), this assertion conveys the information that a sentence s_2 is really true—but then there is, necessarily, another way to convey that very piece of information: we may simply assert s_2. Since s_1 is an arbitrary fictive sentence, this result generalises: if a fictive sentence piggybacks truthful information, then it cannot be impossible to convey this information by other means.

At this point it is clear that Everett contradicted himself: he claimed that fictive sentences piggyback truthful information, and also that the truthful information some of these sentences piggyback cannot be conveyed otherwise. Trivially, in order to eliminate this contradiction, Everett must renounce one of the aforestated claims. Hence, Everett faces a dilemma: he must either deny that fictive sentences piggyback truthful information, or admit that it is possible to convey the truthful information that fictive sentences piggyback by means other than piggybacking. Each horn, I will show, leads to a result that is undesirable for Everett.

Let us start by considering the first one: some fictive sentences do not piggyback information. Let s be one of these sentences. By definition, a fictive sentence is a sentence that involves purported reference to *ficta*, or existential quantification over *ficta*. Since *ficta* do not exist, s is untrue. By assumption, though, s appears to us to be true (cf. the first paragraph of §1), so we may rationally put it forward. Nevertheless, Everett

proposes that we would never put it forward sincerely: on his view, fictive sentences are not believed, but make-believed, and, accordingly, they are always asserted pretencefully. On the face of this proposal, we could justifiably protest that s seems to us to be true and, consequently, that we genuinely believe it, and assert it sincerely. Everett's chief response to this fair protest, we have seen, consists in contending that ordinary utterances of fictive sentences appear to us to be sincere assertions of true statements because the utterances in question serve the serious purpose of conveying pieces of truthful information that cannot be conveyed by other means. These pieces of information, Everett maintains, are conveyed by piggybacking. However, s, by assumption, does not piggyback information; and, if it does not, then Everett cannot satisfactorily explain away our intuitions that assertions of s are truthful and genuine. He must thus admit that, on his view, the assertions in question are wholly empty of cognitive significance. If this is the case, then it is irrational to put s forward, and we are rationally compelled to abandon it. This revisionary result, I believe, is undesirable for Everett: if he endorses it, he has failed the fictionalist purpose of making anti-realism cohere with anti-revisionism.

Let us, then, consider the second horn of the dilemma: the pieces of truthful information that fictive sentences piggyback may be conveyed by means other than piggybacking. *Prima facie*, this horn does not appear problematic: earlier I established that, if a sentence piggybacks a piece of truthful information, then it is possible to convey that very piece of information by means other than piggybacking. Nevertheless, as Everett himself recognises (2005; pp. 624, 647; 2013, p. 54), some fictive sentences are such that it appears hard, or even impossible, to convey by alternative means the information that they piggyback. Suppose that I write a fictional story tonight about an extraordinary creature called 'a'. Then I ask all of my friends to write stories portraying nothing other than creatures modelled upon a. My friends then ask their own friends to write stories portraying nothing other than a-modelled creatures and, thus, a little literary tradition originates where every fictional entity is modelled upon a. Let us call this tradition 'T'. Relative to T, the following fictive sentence is intuitively true:

10. a is a fictional entity such that all the other fictional entities are modelled upon it.

Everett would say that (10) is untrue on the grounds that *ficta*, on his view, do not exist. He would say, moreover, that ordinary assertions of (10) are pretenceful and convey

truthful information at the level of pragmatics. By (Piggybacking), that which ordinary assertions of (10) convey is that a certain sentence s is true in the real world—or, in semantically deflated terms: that which ordinary assertions of (10) express is that, in the real world, s. Consequently, in order to express that which (10) piggybacks, we may just as well assert s—but what sentence, exactly, is s? If we find it, we will have demonstrated that we need not assert (10) in order to express that which ordinary assertions of (10) express: we will have demonstrated that we may just as well assert s instead. In what follows, though, I will argue that it is truly difficult to devise what sentence s might be. If s cannot be found, then the thesis cannot be upheld that, for every fictive sentence, the piece of truthful information that this sentence conveys by piggybacking may be conveyed by alternative means. Thus Everett will be forced back onto the first horn of the dilemma.

s is characterised by two special qualities. The first is that of being ontologically conservative: s does not involve purported reference to *ficta*, nor quantification over *ficta*. If it did, it would be untrue, for *ficta* do not exist. s, however, is not untrue: if ordinary assertions of (10) convey truthful information, and they convey that s is true, then s is true.

The second special quality that s possesses is that of being an adequate paraphrase of (10). There exist multiple competing accounts of paraphrase adequacy, and I shall not assess them here. Instead, I will adopt Quine's account, and justify my choice later. Quine rejects semantic sameness and necessary equivalence as criteria for paraphrase adequacy. His account, instead, is functional: on his view, searching for an adequate paraphrase of a sentence s_1 requires us to pinpoint exactly what is that which we wish to express when we utter s_1, and then find another sentence, s_2, which would allow us to express the same thing (1960, p. 259). In sum, then, s_2 is an adequate paraphrase of s_1 just in case we would be happy to put s_2 forward in place of s_1.

I submit that Everett is bound to accept that s, the sentence that (10) conveys, is a Quinean paraphrase of (10). Suppose that we want to use (10) to describe the real world—and, more specifically, tradition T. Suppose, moreover, that we would not use s in place of (10) to say what we want to say. If we would not use s in place of (10), then it ought to be the case that s does not express what we wish to express by putting (10) forward. Nevertheless, on Everett's view, assertions of (10) are world-oriented pretenceful assertions: they serve the purpose of conveying truthful information at the

level of pragmatics. By assumption, the particular piece of information that assertions of (10) convey is that s is true in the real world—but, also by assumption, s does not express what we want to express. It follows that, on Everett's view, ordinary utterances of (10) convey truthful information that is not the information that we want to convey when we utter (10). If this is the case, then it is irrational to continue using (10) as we do; we must, in other words, revise our use of (10). Everett, however, promised us a theory that would allow us to preserve our ordinary use of fictive sentences. Unless he is willing to admit that his theory is, by his own lights, unsuccessful, he must reject the assumption that led us to conclude that we ought to revise our use of (10). He must reject, therefore, the supposition that s is such that we would not assert it in place of (10). Equivalently: he must admit that s is such that we *would* assert it in place of (10).

Thus I have established that s is a Quinean paraphrase of (10)—even by Everett's own lights. I have also established that s is ontologically conservative. But how, exactly, do we go about searching for s? I suggest that we bear two things in mind when we search for s.

i. (10) involves purported reference to fictional entities, and quantification over fictional entities. In order to construe an ontologically conservative paraphrase of (10), we must paraphrase away all of these committing locutions. In order to do so, we may resort to the fiction operator. Quantification over fictional entities within the scope of the fiction operator is wholly non-committing: I do not commit to fictional entities when I say, for instance, that, according to Julio Cortázar's *Stories of Cronopios and Famas* (1962), there are *cronopios*. Additionally, in footnote 8, I suggested that we assume, for the sake of Everett's argument, that a parafictive statement may be true even if an empty fictive name occurs within the scope of its fiction operator. Under this assumption, an occurrence of a fictive name within the scope of the fiction operator is non-committing. In sum, then: neither quantificational nor referential mentions of fictional entities are committing when they occur within the scope of the fiction operator. In order to eliminate undesirable commitments from (10), therefore, we may try to corner all committing locutions to the scope of this operator.

ii. Tentatively, we may understand that modelling a fictional entity b upon another one c consists in constructing b as bearing a significant number of

the properties that characterise c.

I suggest, on the basis of (i) and (ii), that we consider the following s-candidate:

11.   There is a work of fiction F_1 such that, according to F_1, (an individual a
        exists and it possesses properties $G_1...G_n$). However, 'a exists' is not true in
        reality. Moreover, every work of fiction F is such that, according to F, (for
        every x, (if $x \neq a$, then x possesses a significant number of the members of
        $\{G_1...G_n\}$)).

(11) is ontologically conservative: there is no purported reference to *ficta*, nor quantifica-
tion over *ficta*, outside of the scope of the fiction operator. Its first half, moreover, seems
to satisfactorily express the thought that a is a fictional entity. Nevertheless, there are
three disjoint arguments for the conclusion that (11) is ultimately not a Quinean para-
phrase of (10). The first is this: (11) says that all fictions state that every individual
distinct from a possesses a significant number of the members of $\{G_1...G_n\}$. (11), then,
entails that a figures in all fictions—but (10) entails no such thing. Moreover, there is
nothing that justifies taking an ordinary assertion of (10) as implicating that a figures
in all fictions. Not even at the level of pragmatics, then, does (10) implicate that which
(11) entails. Consequently, if we asserted (11) in place of (10), we would say something
that we originally did not want to say, and we might not wish to do that.

Recall that we are reading (10) as concerning the literary tradition that I called 'T'.
Suppose that we know of a fiction of T that does not feature a. Under this assumption,
we speak truthfully when we put (10) forward, but we would knowingly speak falsely if
we uttered (11) instead. We thus rationally refuse to assert (11) in place of (10)—and
so we ought to conclude that (11) is not a Quinean paraphrase of (10).

Let us now consider (12):

12.   There is a work of fiction F_1 such that, according to F_1, (an individual a
        exists and it possesses properties $G_1...G_n$). However, 'a exists' is not true
        in reality. Moreover, for every fiction F and every entity x, (if $x \neq a$, then,
        (if according to F, (x exists), then according to F, (x possesses a significant
        number of the members of $\{G_1...G_n\}$)))).

The first half of (12), like that of (11), seems to express satisfactorily the thought that
a is a fictional entity. Unlike (11), moreover, (12) does not entail that a figures in every

31

fiction. Sadly, (12) is not ontologically conservative. (12) says that, for every fiction F, and every individual x distinct from a, if x exists in F, then, in F, x possesses a significant number of the members of $\{G_1...G_n\}$. But then one of the occurrences of the fictive name 'a' in (12) does not fall within the scope of the fiction operator, and it is, therefore, committing. (12), in sum, is of no use to us. Let us return to (11).

According to an influential proposal due to Lewis, (1978, p. 42), interpreting a story correctly requires us to assume that the world that story portrays resembles our world in all respects other than those in which the story explicitly describes it as different. On this view, it is interpretatively wrong to imagine, for instance, that persons in the world of Borges's story have three eyes instead of two; and, conversely, we ought to imagine that, in the world in question, Uruguay, for example, is as it is in our world: it is a country located in South America, and South America, in turn, is the Southern portion of one of the continents that exist on Planet Earth—and we ought to imagine that all of this is the case even if the stories do not say so explicitly. If Lewis is right, then the world of Borges's story, like our world, is populated by thousands of persons. Most of these persons, though, do not figure in the stories at all. Hence, the fictional entities that figure in a work of fiction are not necessarily all of the entities that exist in the corresponding fictional world. There is, therefore, a distinction between the entities that exist in a fictional world, and those that are portrayed in the corresponding fiction.

This distinction constitutes the grounds of a second argument to the effect that (11) is not a Quinean paraphrase of (10). (11) says that every fiction states that all the entities that exist and are distinct from a bear a significant number of the members of $\{G_1...G_n\}$. Intuitively, though, (10) does not say that: (10) concerns not the totality of entities that exist in every fictional world, but the totality of fictional entities that are portrayed in one work of fiction or another.

(10), we have seen, concerns tradition T. Suppose that we know that a fiction of T is such that its constitutive sentences do not state, entail nor implicate in any way that all of the individuals that exist possess a significant number of the members of $\{G_1...G_n\}$. Under this assumption, we speak truthfully if we put (10) forward, but we would knowingly speak falsely if we uttered (11). We are thus happy to utter (10), but not (11). It follows, again, that (11) is not a Quinean paraphrase of (10).

(13), unlike (11), is answerable to the distinction between the fictional entities that are portrayed in a work of fiction and the entities that exist in the corresponding fictional

world:

13.   There is a work of fiction F_1 such that, according to F_1, (an individual a exists and it possesses properties $G_1...G_n$). However, 'a exists' is not true in reality. Moreover, for every fiction F and every entity x, (if x is portrayed in F, and, according to F, ($x \neq a$), then, according to F, (x possesses a significant number of the members of $\{G_1...G_n\}$)).

Sadly, (13) is not s. Some of the fictional entities that are portrayed in some works of fiction resemble real entities in such a way that they appear to be numerically identical to the real entities that they resemble. *Funes, the Memorious*, for instance, portrays a town called 'Fray Bentos' that resembles Fray Bentos in such a way that it appears to be numerically identical to Fray Bentos. Whether or not these entities are, indeed, numerically identical is philosophically controversial, but Everett (2013, pp. 63–74) is committed to the thesis that they are.[11] Consequently, on Everett's view, some of the entities that are portrayed in some works of fiction are not fictional, but real. There is, therefore, a distinction between the entities portrayed in a work of fiction, and the *fictional* entities portrayed in that work of fiction. If this is the case, then (13) is not a Quinean paraphrase of (10). (13) says that, for every fiction F, and *for every individual x in our domain of quantification,* if x is portrayed in F, and F says that x is distinct from a, then, in F, x possesses a significant number of the members of $\{G_1...G_n\}$. Thus it is clear that (13) concerns the individuals that exist within our domain of quantification. If it does, though, then it does not concern fictional entities at all: if an individual exists within our domain of quantification, then, trivially, it exists; and, on Everett's view, fictional entities do not exist. (10), by contrast, explicitly concerns fictional entities. If we put (10) forward, we would do so in order to say something about certain fictional entities. Clearly, we cannot use (13) for that purpose. We would, therefore, not assert (13) in place of (10), so (13) is not a Quinean paraphrase of (10).

What we need in order to turn (13) into a Quinean paraphrase of (10), it seems, is to clarify that the entities that possess a significant number of the members of $\{G_1...G_n\}$ are not real, but fictional. Consider, then, the following statement:

14.   There is a work of fiction F_1 such that, according to F_1, (an individual a

---

11. Wellek and Warren (1956) and Riffaterre (1990), on the other hand, oppose this thesis.

exists and it possesses properties $P_1...P_n$). However, 'a exists' is not true in reality. Moreover, for every work of fiction F and every entity x, (if x is portrayed in F, but x does not exist in reality, and, according to F, ($x \neq a$), then, according to F, (x possesses a significant number of the members of $\{G_1...G_n\}$)).

The latter half of (14) is vacuously true: trivially, no individual x in our domain of quantification is such that x does not exist. The latter half of (14), then, is trivially true. Consequently, we convey no valuable information by uttering the latter half of (14); we fail to say anything beyond that which is knowable *a priori*. The latter half of (10), by contrast, is not trivially true, and whatever it is that we pragmatically convey when we put it forward is not trivially true either. The latter half of (10), in sum, does convey valuable information. Consequently, in certain circumstances, we may wish to utter (10), but not (14). It follows that (14) is not a Quinean paraphrase of (10).

Earlier I said that there are three disjoint arguments to the effect that (11) is not a Quinean paraphrase of (10). I have already presented two of them, but not yet the third one. Thus far, I have been assuming that the modelling predicate is adequately paraphrased in terms of the possession of certain qualities (cf. (*ii*) above). This assumption cannot be right, however. Suppose that F_2 is a story where an individual b exists, and possesses a significant number of the members of $\{G_1...G_n\}$. The author of F_2, though, has never read F_1, nor has she ever heard about a. Intuitively, it would be blatantly incorrect to assert that b was modelled upon a: b may share a significant number of properties with a, but, given that the author of F_2 did not intend b to resemble a, we cannot claim that a served as a model for b. Modelling a fictional entity upon another fictional entity, then, requires an intention of the relevant author to portray the former as resembling the latter. If this is correct, then an adequate paraphrase of (10) must include a clause stating something to the effect that all fictional entities other than a were intentionally created to resemble a. Consider, for instance:

15. Every work of fiction F is such that, (according to F, (for every x, (if $x \neq a$, then the author of F intended x to resemble a))).

(15) entails that, for every fiction F, a and the author of F figure in F, and this is not entailed by (10). Moreover, there is nothing that justifies taking an ordinary assertion of (10) as implicating that a figures in every fiction, nor that the author of every work of fiction figures in her own story. Not even by piggybacking, then, does (10) implicate

34

that which (15) entails. Consequently, if we asserted, in place of (10), a paraphrase of (10) involving (15), we would say something that we originally did not want to say—and we might not wish to do that. Suppose that we know of a T-fiction that does not feature a nor its own author. This fiction, clearly, constitutes a counterexample to (15). Hence, in this scenario, we speak truthfully when we put (10) forward, but we would knowingly speak falsely if we uttered, instead, a paraphrase of (10) involving (15). We could, then, rationally refuse to assert, in place of (10), any paraphrase of (10) involving (15). And so we ought to conclude no paraphrase of (10) involving (15) is a Quinean paraphrase of (10).

As an alternative to (15), we could consider the following:

16.   For every work of fiction F and every individual x, (if F is such that, according to F, (x exists and $x \neq a$), then the author of F intended x to resemble a).

Nevertheless, the one instance of the fictive name 'a' that occurs in (16) falls outside of the fiction operator, and so (16) is not ontologically conservative. It is not, therefore, the sentence s that we wish to find.

There are yet more problems than I have pointed out with (11)-(16). Some of these sentences, for instance, speak of fictional entities that are portrayed in stories, and I do not believe that this is ontologically conservative. It might appear that saying that an entity c is portrayed in a certain work of fiction is just to say that, according to the story in question, c exists—but I have already shown that this cannot be the case: there is a distinction, I noted, between the entities that exist in a fictional world, and those that are portrayed in the corresponding work of fiction. Parafictive statements state that there is a binary relation between a work of fiction and a proposition. Consider (5), for instance: it states that a binary relation exists between *Funes, the Memorious* and the proposition that Ireneo is a hypermnesiac. Some accounts of propositions allow this proposition to exist even if Ireneo does not (see, e.g., Braun, 1993; 2005; Friend, 2011; 2014), so (5) and the like may well be ontologically conservative. Nevertheless, it is truly very hard to see how one could construe as ontologically conservative those sentences that state that a fictional entity is portrayed in a certain work of fiction, for these sentences affirm not that there is a binary relation between a story and a proposition, but that there is a binary relation between a work of fiction and a fictional entity. It seems to me to be evident that no binary relation can exist between an object

35

and nothing else,[12] so the existence of a binary relation between a story and a *fictum* requires that the *fictum* in question exists. In sum, therefore: a sentence that states that a *fictum* is portrayed in a work of fiction is not conservative.

I hope I have now shown that it is not at all easy to find an ontologically conservative, Quinean paraphrase of (10). I argued that there are three reasons, independent of one another, why (11) does not constitute such a paraphrase; (11), in other words, is triply flawed for our purposes. (12)-(16) are five different attempts to fix these flaws, but they, I have proved, are flawed, too. To my mind, the failure of (11)-(16) as ontologically conservative, Quinean paraphrases of (10) should suffice to demonstrate that the prospects for finding one such paraphrase are rather disheartening; and, moreover, that it is probably the case that the piece of information that ordinary assertions of (10) convey cannot be conveyed by means other than putting (10) forward.

The second horn of the dilemma that I posed for Everett at the beginning of this section, consequently, leads to a daunting challenge: on Everett's view, ordinary utterances of (10) piggyback that a sentence s is true in the real world; I challenge Everett to show us what sentence, exactly, s is. If he succeeds, then we will be able to believe him that there is a piece of information that ordinary assertions of (10) piggyback. If he fails, though, he will be forced back into the first horn of the dilemma: he will be forced to admit that some fictive sentences do not piggyback information. And this horn, I have shown, quickly leads to the conclusion that Everett's theory is unsuccessful.

In order to evade this challenge, Everett could repudiate my contention that s, the sentence piggybacked by (10), is a Quinean paraphrase of (10). Clearly, though, s must be a paraphrase of some kind of (10); it must, in some sense or another, say that which (10) says: otherwise, we would not be able to believe Everett's thesis that our assertions of (10) express that s is really true. As I mentioned earlier, there are multiple competing accounts of paraphrase adequacy; but, as it happens, Quine's appears to be remarkably flexible amongst its alternatives (see, e.g., Von Solodkoff, 2014b, p. 576). My requirement that s is an adequate paraphrase of (10), then, can be hardly thought to be unduly stringent; and it cannot be considered arbitrary either: I chose an especially flexible account of paraphrase adequacy precisely in order to increase our chances of finding an adequate paraphrase of (10). Finally, and more importantly, I showed earlier that Everett is bound to accept that s is a Quinean paraphrase of (10). I submit,

---

12. In connection with this, see Button (2012).

consequently, that Everett cannot escape my challenge by rejecting my Quinean demand upon s.

If my arguments in this section are sound, then Everett faces an unpalatable dilemma. One of its horns leads to the conclusion that he has failed in the fictionalist project of making anti-realism cohere with anti-revisionism; the other leads to a challenge that, to me at least, seems rather daunting. In any case, therefore, Everett is in a precarious position.

5   Conclusion

In this chapter, I advanced two objections and a dilemma against Everett's theory. If my arguments are sound, then this theory is not satisfactory. In the introduction of this chapter, I noted that Everett's theory is prominent amongst the various fictionalist theories that exist of fictive statements. By arguing that it is unsatisfactory, I hope to have established that the prospects for success of the project of fictive fictionalism are disheartening.

In light of the above, I will devote the next five chapters to the principal rival of fictive fictionalism: fictional realism. As I said in the introduction of this chapter, fictive fictionalists aim at honouring our use of fictive sentences without postulating *ficta*. Fictional realists, by contrast, contend that honouring our use of these sentences *requires* us to posit *ficta*. Once fictional entities are posited, though, the question arises of what ontological category they pertain to. Different fictional realists offer different answers to these questions, and, therefore, there are various fictional-realist theories which comprise different accounts of the ontology of *ficta*. In what follows, I will discuss two of these theories, and I will advance a novel one. Let us see.

Chapter Two:
Creationism and Reference

Creationism is the view that fictional entities are abstract artefacts created by writers (see, e.g., Van Inwagen, 1977; 1983; 2000; Salmon, 1998; Thomasson, 1998; Voltolini, 2006; Kripke, 2013).[13]

Prima facie, this view is frankly disconcerting: it seems platitudinous that there are no fictional entities; that, for every name '*a*', the statement that *a* is fictional invariably sanctions the conclusion that *a* does not exist.[14] There is, though, a powerful rationale behind the postulation of *ficta*. Let us see.

Let the term 'fictive sentences' designate the sentences that involve reference to *ficta*, or existential quantification over *ficta*. Some fictive sentences are intuitively true. Witness:

1. Ireneo is a hypermnesiac.
2. Ireneo was created by Jorge Luis Borges in 1942.
3. Some fictional entities are more realistic than others.

---

13. If there exists a characterisation of abstract entities which deserves to be called 'standard', it is the following: abstract entities are entities that are not located in spacetime (Falguera, Martínez-Vidal & Rosen, 2022, §3.5). Creationists, however, cannot endorse this characterisation, for they believe that *ficta* are abstract, and that they were brought into existence by writers at determinate points in time. The question arises, therefore, of what, on the creationists' view, characterises the entities that they classify as abstract; what, in other words, is the property in virtue of which these entities deserve the classification in question. Perhaps creationists could claim that, while some *abstracta* (e.g., *ficta*) are located in time, the totality of abstract entities possess the property of being non-spatial. Nevertheless, the theory of relativity states that space and time constitute a non-separable manifold. If this is statement is true, then it appears that the thought is absurd that some entities are located in time, but not in space (Falguera, Martínez-Vidal & Rosen, 2022, §3.5). If this is the case, then creationists face the challenge of explaining this apparent absurdity away, or formulating an alternative, non-spatiotemporal characterisation of *abstracta*.

14. See Thomasson (1998, pp. 12–13) and Von Solodkoff (2014) for two different proposals on how the creationist may honour our intuitions that fictional entities do not exist.

If (1)–(3) are true indeed, then the name 'Ireneo' refers, and the existential quantifier in (3) ranges over a domain that comprises fictional entities. It appears, therefore, that our commitment to the truth of (1)–(3) compels us to admit fictional entities into our ontology.[15]

In sum: those who posit fictional entities do so with the purpose of honouring our intuitions that some fictive sentences are true. In this chapter, however, I will argue that creationists have not yet attained this aim.

Fictive sentences belong to two distinct kinds: the kind of fictional sentences and the kind of critical sentences. Fictional sentences are sentences where fictional entities are attributed their fictional properties; critical sentences are sentences where fictional entities are attributed critical properties. For every *fictum a*, *a*'s fictional properties are the properties that are attributed to *a* in the story where *a* figures. Ireneo's fictional properties, for instance, are, amongst others, the following: being a flesh-and-blood man; being Uruguayan; being a hypermnesiac; and suffering from insomnia. (1), then, is a fictional sentence: it is a sentence where a fictional entity is attributed one of its fictional properties. *a*'s critical properties, on the other hand, are the properties that we attribute to *a* when we observe it from the perspective of the literary critic—when we regard it, that is to say, as an authorial creation that may be aesthetically judged. The following are examples of critical properties: having been created by an author at a certain point in time; being well constructed; being realistic; being revolutionary; and being typical or exemplar of a certain literary tradition. (2) and (3), then, are critical sentences: they are sentences where fictional entities are attributed critical properties.

Fictional entities, within the creationist framework, instantiate their critical properties, but not their fictional properties. Consider Ireneo, for example. On the creationist's view, Ireneo is an abstract entity produced by Borges's authorial endeavours; it is, therefore, an authorial creation that instantiates certain aesthetic qualities. If Ireneo is an abstract artefact, though, then he does not instantiate any one of his fictional properties: evidently, an abstract artefact cannot instantiate the property of being a flesh-and-blood, Uruguayan man; it cannot be a hypermnesiac, and it cannot suffer

---

15. Once fictional entities have been admitted into our ontology, the question arises of what their ontological nature is. Some (e.g., Parsons, 1980; Zalta, 1983; Jacquette, 1996) believe that fictional entities are Meinongian non-existents; others (e.g., Lewis, 1978; Stone, 2010; Haraldsen, 2017; 2020), that they are concretist non-actual individuals. Nowadays, however, the most popular view seems to be creationism: the view that they are abstract artefacts created by writers.

from insomnia.

In more general terms: *ficta* in general are portrayed in their stories as bearing properties that *abstracta* cannot possess: they are depicted as spatiotemporal beings, as living creatures, as having thoughts and emotions, and as capable of speaking and performing all kinds of deeds. Thus, within the creationist framework, *ficta* in general are such that they cannot instantiate their fictional properties.

If fictional entities within the creationist framework instantiate their critical properties, but not their fictional properties, then, trivially, critical sentences are true within the creationist framework, but not fictional sentences. This result is severely problematic, for we utter fictional sentences often and, intuitively, we speak truthfully when we do so. Intuitively, for instance, we speak truthfully when we say that Ireneo is incapable of oblivion, and that he learned Latin in a matter of days—but none of that is true within the creationist framework. The creationist ontology, in sum, is in conflict with our discursive practices.

Recently, Von Solodkoff and Woodward (2017) proposed an elaborate solution to this conflict drawing from the solutions that other creationists proposed in the past (i.e., Van Inwagen, 1977; 1983; Salmon, 1998). In what follows, I will argue that the proposal in question fails. In §1, I will reconstruct the creationists' proposal. In §2, I will argue that it fails. And, finally, in §3 I will address two rejoinders to my argument which could be raised on behalf of the creationist.

1   Holding

The creationist believes that Ireneo is an abstract artefact. If he is an abstract artefact, then he is not a hypermnesiac. If he is not a hypermnesiac, then (1) is untrue. (1), however, is intuitively true, so the creationist doctrine violates our intuitions.

The creationist admits that Ireneo cannot instantiate the property of being a hypermnesiac if he is abstract. Nevertheless, she claims, he undoubtedly bears to this property a relation that he does not bear to, say, the property of being invisible. Even if fictional entities cannot instantiate their fictional properties, she concludes, there must exist a relation that relates them to these properties.[16] The creationist calls this relation 'hold-

---

16. This argument is originally due to Van Inwagen (1983, p. 75).

ing'.[17] The creationist, moreover, contends that fictional sentences are ambiguous: they can be read as stating that there is a relation of instantiation between fictional entities and their fictional properties, or that there is a relation of holding between them. (1), for instance, is ambiguous between the following two claims:

4.   Ireneo instantiates the property of being a hypermnesiac.
5.   Ireneo holds the property of being a hypermnesiac.

(4), the creationists says, is false, but (5) is true, and the reason why we intuitively think that (1) is true is that we use it to express (5) rather than (4) (see, e.g., Van Inwagen, 1983, pp. 75–76). Thus, the creationist concludes, our intuitions that (1) is true have been done justice.

The creationist believes that *ficta* instantiate properties of several kinds. They instantiate, first, logical or high-category properties like existence and self-identity (Van Inwagen, 2000, p. 245–246). They also instantiate critical properties (Van Inwagen, 2000, p. 245–246). And, finally, they instantiate the property of being abstract. The properties that *ficta* hold, the creationist maintains, are properties like "being human[,] being a resident of Hannibal, Missouri, being an orphan who has a mysterious benefactor, [and] being a witch" (Van Inwagen, 2000, p. 246). All of these properties are attributed to fictional entities in the stories where they figure. This clearly suggests that the relation of holding is to be analysed as follows:

Holding
A *fictum* a holds the property of being $G \leftrightarrow$ according to the relevant story, a is G.

Van Inwagen however, repudiated this analysis, and claimed that the relation of holding is unanalysable: "What is this holding? I cannot define it. I can only give examples. But we all understand it well enough" (1983, p. 75). Von Solodkoff and Woodward (2017, pp. 411–412) believe that the reason why Van Inwagen repudiated (Holding) is that he could not make sense of its right-hand side: on his view, no propositions whatsoever are true according to works of fiction because the sentences that constitute these works "do not represent attempts at reference or description" (1977, p. 301). Nevertheless, Von Solodkoff and Woodward (2017, pp. 412–414) follow Salmon (1998, p. 297) in contending that this hypothesis is unacceptable: it renders "problematic and

---

17. It was Van Inwagen (1983, p. 75) who originally coined this term.

mysterious" the truth of incontrovertibly true sentences like the following:

6.   According to the story, Ireneo is a hypermnesiac.

In what follows, I shall follow Salmon, Von Solodkoff and Woodward, and assume, without further argument, that Van Inwagen's doctrine is unacceptable. If it is unacceptable, then it seems, for the reasons that I adduced above, that the relation of holding is to be analysed in terms of (Holding). In the following section, however, I will show that (Holding), too, is unacceptable.

2   Against Holding

(Holding) suggests a peculiar picture of what fiction is. As I said earlier, the creationist thinks that (1) is ordinarily used to express (5); and, according to (Holding), (5) is true just in case (6) is true. If the name 'Ireneo' in (5) refers to an abstract object, as creationists maintain, then, by (Holding), it refers to an abstract object in (6) too. But then (Holding) entails that the *abstractum* Ireneo holds the property of being a hypermnesiac just in case Borges's story says *of it* that it is a hypermnesiac. If it does, however, then it is significantly different from what we intuitively thought that it is: intuitively, Borges's story concerns a man who does not really exist, but now it appears that it concerns, instead, an abstract artefact: it tells us, not of a non-existent man, but of an abstract object, that it is a hypermnesiac.

Salmon, Von Solodkoff and Woodward are willing to bite this bullet. Borges, they claim, did indeed write a story that says, of a certain *abstractum*, that it is a Uruguayan, hypermnesiac man—but Borges did not thereby commit a silly mistake (Salmon, 1998, pp. 300–301, 316; Von Solodkoff and Woodward, 2017, p. 414). On their view, it appears, what Borges did was very much akin to that which children do when they play make-believe games, and say pretencefully of a wooden stick that it is a sword, or of a doll that it is a baby. Borges, that is to say, played a make-believe game with a number of abstract objects to which he pretencefully ascribed various properties in the act of creating a story. Salmon exhorts us not to fixate on the fact that *ficta* are abstract and, thus, not at all like sticks or dolls (1998, p. 301): seen in the proper light, he says, to pretend that an *abstractum* is a hypermnesiac man is no stranger than pretending that Marlon Brando is Don Corleone (1998, p. 316). To my mind, therefore, the account of fiction that Salmon, Von Solodkoff and Woodward advocate for is one where *ficta*, and the actors and the props on the stage of a dramatic production are

42

very much alike: in the act of engaging with a work of fiction, we imagine them all as bearing properties that they do not bear in reality: we pretend that Laurence Olivier is the Prince of Denmark, and we pretend that a wooden artefact is a poisoned sword; and, analogously, we pretend that an abstract object is a hypermnesiac man. The only difference between actors and props on the one hand, and *ficta* on the other, is that the latter are abstract—but that, Salmon insists, is unimportant.

I believe that this pretence-theoretic account of fiction is unconvincing. Halldór Laxness's *Lilja* is the story of an Icelandic man called 'Nebúkadnessar Nebúkadnessarsson', and his friendship with a child whose name is also the name of this story. At one point, Nebúkadnessar is described by the story's wholly reliable narrator as uttering the following sentence sincerely: 'Lilja is fond of raisins'. Lilja's fondness for raisins is never mentioned again, but the following sentence is intuitively true nonetheless:

7.   Lilja is fond of raisins.

Since abstract objects cannot instantiate the property of being fond of raisins, (7) must be read as saying that the *abstractum* Lilja holds the property in question. According to (Holding), this is true just in case the *abstractum* in question is said in its story to be fond of raisins—but is it? I think that it is not. Salmon, Von Solodkoff and Woodward would say that Laxness used the name 'Lilja' to refer to an *abstractum* that he, within a certain pretence, imagined as a child that loves raisins. And perhaps they are right, but that is irrelevant now. In the story, it is Nebúkadnessar who says that Lilja is fond of raisins, and Nebúkadnessar himself is a fictional entity. He is, then, an abstract object that cannot speak. Clearly, what happened, on the creationist picture, is that Laxness pretencefully attributed to the *abstractum* Nebúkadnessar a certain property: the property of being a man who utters the sentence 'Lilja is fond of raisins'. Let us, though, stop for a moment to investigate what exactly this property is. Laxness may have been caught in the pretence that certain *abstracta* are flesh-and-blood humans; but, in the story, Nebúkadnessar is not caught in any such pretence. It is clear in the story that Nebúkadnessar's assertion was not pretenceful, but genuine. Now, what is that which Nebúkadnessar asserted? What is the proposition expressed by 'Lilja is fond of raisins', as asserted by Nebúkadnessar? Evidently, it cannot be the proposition that the abstract object Lilja is fond of raisins: if it were, then, given that we have established that Nebúkadnessar was speaking sincerely, we would have to conclude that he was seriously confused—but he was not: the story does not portray him as being

43

confused in any way. It follows that the property that Laxness pretencefully attributed to Nebúkadnessar is *not* that of having said of an abstract object that it is fond of raisins; that, unsurprisingly, we are *not* supposed to imagine Nebúkadnessar as saying, of an abstract object, that it is fond of raisins.

Imagine, for the sake of analogy, an actor on the stage of a dramatic production of the medieval Scottish ballad *Edward*. The actor who is playing Edward points to a wooden artefact and says: 'This is the sword that killed my father'. In the fiction, Edward genuinely believes that the object that killed his father is a sword, not a wooden artefact, and he is definitely in his right mind. Clearly, then, Edward is not supposed to be imagined as saying of a mere wooden artefact that it is the sword that killed his father: he would have to be insincere or confused in order to say that. The wooden artefact on stage, then, is a *mere prop:* it is meant to help us attain a depiction in our minds of the scenes that the fiction narrates; it *represents* the thing that Edward is talking about, but it is not itself that thing. Hence, when Edward says 'This is the sword that killed my father', the demonstrative that he uses is not meant to refer to a wooden artefact at all.

Let us now return to *Lilja*. I suggested earlier that, on the creationist proposal, fictional entities function very much like props. If this suggestion is right, then supposing that Nebúkadnessar said of an *abstractum* that it is fond of raisins is as confused as supposing that Edward said of a wooden artefact that it is the sword that killed his father. If this is right, then Laxness's story cannot be interpreted as portraying Nebúkadnessar as saying of an *abstractum* that it is fond of raisins.

If this conclusion is sound, then the *abstractum* Lilja is never said in Laxness's story to be fond of raisins. Thus, according to (Holding), this *abstractum* does not hold the property of being fond of raisins. It follows that (7) is, contrary to our intuitions, not true. There are, therefore, intuitively true fictional sentences that cannot be done justice within the creationist framework.

Now, (7) describes a *fictum* as another *fictum* described it in the story where they both figure. Importantly, though, the fictional sentences that the creationist struggles to honour are not all like (7). Let us see.

Lilja is narrated by Nebúkadnessar's neighbour, whom we have every reason to believe is a reliable narrator. At one point he says that Nebúkadnessar often wears a bowler

hat, and this is never mentioned again. The following sentence, however, is intuitively true:

8.   Nebúkadnessar often wears a bowler hat.

Creationists believe that 'Nebúkadnessar' in (8) refers to an abstract object. Given that abstract objects cannot wear hats, creationists must read (8) as saying that the *abstractum* Nebúkadnessar holds the property of wearing a bowler hat. According to (Holding), this is true just in case the *abstractum* Nebúkadnessar is said in the story to wear a bowler hat; but it seems to me that it is not. Nebúkadnessar's neighbour is a fictional entity and, thus, an abstract object that cannot speak. Laxness pretencefully attributed to it the property of having uttered the sentence 'Nebúkadnessar often wears a bowler hat'; but the occurrence of the name 'Nebúkadnessar' within this utterance cannot be understood as referring to an abstract object. In the story, Nebúkadnessar's neighbour was not engaged in the pretence that *abstracta* are flesh-and-blood humans, nor was he confused when he described Nebúkadnessar. Clearly, then, we are not meant to imagine him as saying of an abstract object that it often wears a bowler hat. It follows that the proposition expressed by his utterance was not the proposition that the *abstractum* Nebúkadnessar often wears a bowler hat. If this is right, then the *abstractum* Nebúkadnessar is never said in the story to often wear a bowler hat. By (Holding), it follows that (8) is, contrary to our intuitions, not true.

Thus the conclusion has been reached again that there are intuitively true fictional sentences that cannot be done justice within the creationist framework.

3   Objections

In order to try to block the conclusion that I advanced above, the creationist could analyse the relation of holding in slightly different terms:

Holding′
A *fictum* a holds the property of being $G \leftrightarrow$ according to the relevant story, there is an individual called 'a' who is G.

In Laxness's story, the creationist could claim, there is an individual called 'Lilja' who is fond of raisins. By (Holding′), then, the *abstractum* Lilja holds the property of being fond of raisins. (7), then, is true; and, by the same token, (8) is true, too.

I believe that this revised analysis of the relation of holding is unacceptable for two reasons. Towards the end of Laxness's 'Lilja', Nebúkadnessar's neighbour, the narrator of the story, meets an old woman grieving at Nebúkadnessar's funeral. She and Nebúkadnessar met when they were young, she says, but then she got married and moved away to Keflavík. Her name is 'Lilja'. According to Laxness's story, then, there is an individual called 'Lilja' who is an old lady. By (Holding′), it follows that the *abstractum* Lilja that we have been talking about all along does not only hold the property of being fond of raisins, but also the property of being an old lady. Hence, the following sentence is true:

9.   Lilja is fond of raisins and she is an old lady.

This result, of course, is incorrect: as I said earlier, the story says that the individual Lilja who is fond of raisins is a child. Perhaps, though, the creationist could protest that this incorrect result does not in fact follow from (Holding′). The name of the child and the name of the old lady, the creationist could contend, are not really the same: we should conceive of them as bearing distinct elided numerical subscripts. The name of the young protagonist, let us say, is 'Lilja$_1$' whereas the other one is 'Lilja$_2$'. If this is the case, then the *abstractum* corresponding to the child Lilja holds the property of being an old woman just in case, according to the story, there is an individual called 'Lilja$_1$' who is an old woman—which there is not. (Holding′), the creationist concludes, remains undefeated.

Nevertheless, at this point I would like to ask what determines whether an occurrence of 'Lilja' in Laxness's story is an occurrence of 'Lilja$_1$' or 'Lilja$_2$' if none of the occurrences in question contains any subscript whatsoever. The obvious response to this question seems to be that these names are distinguished by the fact that their uses are associated with distinct referents; that, simply put, they refer to distinct things. Unfortunately, this response begs the question that my arguments against creationism are unsound. If the various occurrences of 'Lilja' within the text that constitutes Laxness's story have different referents, then, trivially, they must have referents; and, within the creationist framework, there are no other candidate referents for them than a pair of abstract objects brought into existence by Laxness. What I have been arguing all along, however, is that the occurrences of purportedly referring expressions in fictional stories cannot, in general, be thought of as referring to abstract objects like these. When the narrator of Laxness's story tells us that Nebúkadnessar wears a bowler hat, for instance,

'Nebúkadnessar' does not refer to an *abstractum*. If my arguments are sound, then we are not allowed to suppose that all occurrences of purportedly referring expressions in fictional stories have referents within the creationist framework; and, if this is the case, then there are, it appears, no grounds for us to identify occurrences of 'Lilja' in Laxness's text as being really occurrences of 'Lilja$_1$' or of 'Lilja$_2$'. Consequently, we cannot determine that there is an individual called 'Lilja$_2$' who is an old lady, but not an individual called 'Lilja$_1$' who is an old lady. And, if there is no way for us to do this, then the undesirable result that (9) constitutes cannot be avoided.

Importantly, there is a second and wholly independent reason to reject (Holding′). Laxness's story does not literally say that there is an individual called 'Nebúkadnessar' who often wears a bowler hat: Nebúkadnessar's neighbour merely states 'Nebúkadnessar often wears a bowler hat'. To my mind, it is unclear how that statement is supposed to be used to derive, by means of (Holding′), the conclusion that (8) is true. According to (Holding′), (7) is true just in case there is, according to the story, an individual called 'Nebúkadnessar' who often wears a bowler hat. Now, there are good reasons to believe that fictions are closed under entailment; that, in other words, the following principle holds for every fiction F:

> **Fictive Closure**
> If p is true according to F, and p entails q, then q is true according to F.

I think, however, that (Fictive Closure) does not sanction the use of an existential generalisation to go from 'Nebúkadnessar often wears a bowler hat' to the conclusion that there is an individual called 'Nebúkadnessar' who often wears a bowler hat. According to the arguments that I have advanced, the name 'Nebúkadnessar' in 'Nebúkadnessar often wears a bowler hat' cannot be thought of as referring to an *abstractum*: in the story, Nebúkadnessar's neighbour did not attribute to any *abstractum* the property of wearing any garments. If this is right, then creationists have failed in providing a plausible referent for 'Nebúkadnessar' in the neighbour's utterance of 'Nebúkadnessar often wears a bowler hat'. Thus the name is, at this point, empty. If it is empty, then the inference from the neighbour's 'Nebúkadnessar often wears a bowler hat', to the conclusion that there is an individual called 'Nebúkadnessar' who wears a bowler hat is illegitimate. The creationist, in other words, cannot obtain the required right-hand side of the relevant instance of (Holding′) in order to derive (8). (8), therefore, cannot be regarded as true before creationists address the problem that I have been discussing all

along—the problem, that is to say, of giving us a plausible referent for the occurrences in general of the name 'Nebúkadnessar' within Laxness's story.

There is something else that creationists could adduce at this point. In the introduction of their paper, Von Solodkoff and Woodward explicitly state that the analysis of the relation of holding that they adhere to is (Holding):

> We defend [...] that to say that [Ireneo] holds the property of [being a hypermnesiac] is just to say that the singular proposition [⟨Ireneo, being a hypermnesiac⟩] is true according to [Borges's story] (2017, p. 408).

Afterwards, they seem to weaken somewhat their commitment to (Holding) by claiming that "we might see [(6)] as an analysis of [(5)]" (2017, p. 410). Later on, however, they suggest that the analysis that they favour is ultimately not (Holding) at all, but a different one. At some point, they claim to adhere to a Waltonian account of truth in fiction:

> Following Kendall Walton's (1990) pioneering and influential work, we think that the key to understanding fictionality is the idea that works of fiction prescribe imaginings. The core idea here is that, when we engage with a given work of fiction, we are supposed to imagine that certain things are the case. Exactly how these prescriptions are generated is controversial but we assume that they are generated by the interplay between objective features of the work and features of the interpretative or critical context in which we engage with the work.

> [...] the approach to fictionality that we favour holds that singular propositions like [⟨Ireneo, being a hypermnesiac⟩] and [⟨Ireneo, being Uruguayan⟩] are true according to [Borges's story] insofar as the work prescribes imagining that these propositions are the case (2017, pp. 412–414).

The account of truth in fiction that Von Solodkoff and Woodward endorse, then, is the following:

Waltonian Truth
According to fiction F, $p \leftrightarrow F$ prescribes imagining that proposition p is the case.

After declaring their adherence to (Waltonian Truth), Von Solodkoff and Woodward suggest that the analysis of the relation of holding that they favour is the result of analysing the right-hand side of (Holding) in accordance with (Waltonian Truth):

> Within the context of our preferred Waltonian setting, to say that [Ireneo holds the property of being a hypermnesiac] is to say that [Borges's story] prescribes us to imagine the singular proposition [⟨Ireneo, being a hypermnesiac⟩] (2017, p. 420).

On the face of this passage, it seems that the analysis of the relation of holding that Von Solodkoff and Woodward adhere to is the following:

Holding″

A *fictum a* holds the property of being $G \leftrightarrow$ the relevant fiction prescribes imagining that the proposition Ga is the case.[18]

I believe, however, that, under the creationist assumption that fictional entities are abstract, the Waltonian account of truth in fiction is patently mistaken. Consider our favourite substitution instance of (Waltonian Truth):

10.   According to the story, Ireneo is a hypermnesiac $\leftrightarrow$ this story prescribes imagining the singular proposition ⟨Ireneo, being a hypermnesiac⟩ is the case.

The lefthand side of (10) is (6). (6), we have seen, is intuitively true. If it is true, then the righthand side of (10) is true too. The creationist believes that Ireneo is an abstract object, so they believe that both of the occurrences of 'Ireneo' in (10) refer to an *abstractum*. If the righthand side of (10) is true, then Borges's story prescribes imagining that an *abstractum* is a hypermnesiac—but does it? There is obviously at least one sense in which it does not. If fictions prescribe imaginings, then there are fictions that prescribe imagining that certain humans are the bearers of supernatural powers, and fictions that prescribe imagining that certain animals speak; but Borges's story is definitely not one that prescribes imagining that certain abstract objects are hypermnesiacs. Salmon, however, believes that this is not at all the sense in which creationists think that *abstracta* are imagined as wizards, pirates and whatnot, when we engage with works of fiction. "That abstract entities are human beings is not something we pretend", he says; "but there are abstract entities that we pretend are human beings" (1998, p. 316). There is, that is to say, a distinction between *de dicto* and *de re* readings

---

18. It is important to point out that the proposal in question is not the proposal that fictions prescribe imagining that certain propositions are true. Von Solodkoff and Woodward are very clear that they do not think that fictions prescribe metalinguistic imaginings (2017, pp. 412–414). On their view, that is to say, fictions do not prescribe imagining that certain non-referring names refer, nor that certain untrue propositions are true. This metalinguistic account of the imaginings prescribed by fictions, they argue, is a "poisoned pawn" (2017, p. 412), but I will not discuss their arguments here.

of sentences about what we pretend: on the latter, they are true; on the former, false.

Von Solodkoff and Woodward are aware of this distinction, and endorse it (2017, p. 414). To my mind, however, it is clear that the right-hand side of (Holding″) unambiguously concerns *de dicto* imaginings. A substitution instance of (Waltonian Truth) involving *de re* imaginings of Ireneo and the property of being a hypermnesiac is definitely not (10), but (11):

11.   According to the story, Ireneo is a hypermnesiac ↔ Ireneo is such that this story prescribes imagining that he is a hypermnesiac.

But then the account of truth in fiction that Von Solodkoff and Woodward favour is not (Waltonian Truth), but (Waltonian Truth′):

Waltonian Truth′
According to fiction F, $Ga ↔ a$ is such that F prescribes imagining that it is G.

Accordingly, the analysis of fiction that Von Solodkoff and Woodward should endorse follows by transitivity not from (Holding) and (Waltonian Truth), but from (Holding) and (Waltonian Truth′):

Holding‴
A *fictum* a holds the property of being $G ↔ a$ is such that the relevant fiction prescribes imagining that it is G.

Unfortunately, (Waltonian Truth′) is wrong. In the previous section, I said that *ficta* are very much like actors and props on the creationist view: they are things to which we pretencefully attribute properties in the act of engaging with fiction. Just as we imagine that Marlon Brando is Don Corleone, Salmon said, we imagine that a certain abstract object is a hypermnesiac. Von Solodkoff and Woodward, we have seen, have it that fictions prescribe imaginings. In particular, Borges's story prescribes imagining that Ireneo is a hypermnesiac. By analogy, we must conclude that *The Godfather* prescribes imagining that Marlon Brando is Don Corleone. Of course, this idea requires precisification in accordance with Salmon's insight that there is an important *de re/de dicto* distinction here: it is not the case that *The Godfather* prescribes imagining that Marlon Brando is Don Corleone, but Marlon Brando is such that 'The Godfather' prescribes imagining that he is Don Corleone. That seems plausible enough; but, if it is true, then, by (Waltonian Truth′), it follows that, according to *The Godfather*, Marlon Brando is

Don Corleone—and that is clearly false. (Waltonian Truth′) is, thus, incorrect; and, consequently (Holding‴) is not available to Von Solodkoff and Woodward.[19]

Perhaps at this point creationists would want to claim that the way in which we imagine that Marlon Brando is Don Corleone is not the way in which we imagine that a certain *abstractum* is a hypermnesiac; that, more precisely, it is true that Ireneo is such that Borges's story prescribes imagining that he is a hypermnesiac, but false that Marlon Brando is such that *The Godfather* prescribes imagining that he is Don Corleone. If this is the case, though, then creationists owe us an account of what it is that we do with abstract objects when we pretencefully attribute to them properties like being a hypermnesiac. In §2, I showed, following Salmon, that the idea that we imagine certain *abstracta* as being hypermnesiacs, pirates or witches, as having thoughts and emotions, and experiencing memorable adventures makes sense and appears acceptable *when we compare it with what we do with actors and props.* If it turns out, after all, that this analogy fails, then the burden is on the creationist to explain to us how exactly our engagement with fiction is to be understood: if Ireneo is an abstract artefact, but not a prop, then how is it the case that imagining him as a hypermnesiac is not plainly a silly mistake?

4 Conclusion

As I said in the introduction of this chapter, those who postulate fictional entities do so with the purpose of honouring the intuitive truth of fictive statements. In this chapter, I argued that the creationist has not yet attained this aim: the creationist, I argued, lacks an analysis of the relation of holding which allows her to honour the intuitive truth of fictional sentences. Thus I conclude that creationism is unsatisfactory. In light of this result, the following chapter I will devote to a prominent alternative: possibilism.

---

19. Towards the end of their (2017), Von Solodkoff and Woodward claim that their analysis of holding allows them to dispose of one theoretically weighty creationist doctrine: the thesis that fictional sentences are ambiguous between an instantiation precisification and a holding precisification. With the analysis in question at hand, they contend, we may suppose that fictional sentences are nothing other than ellipses for statements about what is true in the relevant stories. On this view, (1), for instance, is simply an ellipsis for (6). To my mind, however, this contention pressingly poses the question of what the purpose was of Von Solodkoff's and Woodward's lengthy discussion of holding: if fictional sentences are ellipses for according-to-the-story statements, cannot fictionalists account for the truth of these sentences simply by appealing to (Waltonian Truth′)? In any case, I just showed that (Waltonian Truth′) is unacceptable, so my final contention stands: creationists lack a working account of the truth of fictive sentences.

Chapter Three:
Possibilism and the Semantics of Instantial Terms

Fictional realism is the view that there are fictional entities. That fictional entities do not exist appears platitudinous, but fictional realism is robustly motivated. Consider:

1. Ireneo is a hypermnesiac.
2. Ireneo is a Uruguayan man.
3. Ireneo was created by Jorge Luis Borges in 1942.
4. Some fictional entities are more realistic than others.

(1)–(4) are intuitively true. If they are true indeed, then the name 'Ireneo' refers, and the existential quantifier in (4) ranges over a domain that comprises fictional entities. Hence, fictional realists conclude, we must admit *ficta* into our ontology.

The ontological nature of fictional entities is a matter of controversy amongst fictional realists. Some of them (e.g., Van Inwagen, 1977, 1983, 2000; Salmon, 1998; Thomasson, 1998; Voltolini, 2006; Kripke, 2013, pp. 55–101) are creationists: they believe that fictional entities are abstract artefacts created by writers. Some others (e.g, Lewis, 1978; Stone, 2010; Haraldsen, 2017; 2020) are possibilists: they think that fictional entities are concretist *possibilia*—or, equivalently: concretist non-actual individuals. I devoted Chapter Two to creationism; this chapter I will devote to possibilism.

For decades it was thought that possibilism had been defeated by an objection due to Kripke. Recently, however, Caplan (2016, p. 345) and Haraldsen (2017; 2020) contended independently that possibilism may be vindicated by way of an appeal to the theory of arbitrary reference (Breckenridge & Magidor, 2012) (henceforth 'AR'). In this chapter,

I will argue that Caplan and Haraldsen are in error.

In §1, I will present the possibilist theory, and reconstruct Kripke's objection. In §2, I will introduce AR, and explain how Caplan and Haraldsen propose that the possibilist can use it to overcome Kripke's offence. In §3, §4 and §5, I will advance two mutually independent arguments against Caplan and Haraldsen. The first shows that endorsing AR will lead the possibilist astray, and, in particular, that it will lead her to betray her own doctrine. The second states that AR is severely dubious. If my second argument is sound, then the possibilist who endorses AR stands on unstable grounds. I will conclude that the prospects of success for the project of vindicating possibilism are much more disheartening than they appeared.

1   Possibilism and Kripke's Objection

According to Lewis (1986), there exist infinitely many, maximally connected regions of spacetime—infinitely many regions of spacetime, that is to say, that are radically disconnected from one another. These regions of spacetime, Lewis claims, are ontologically on a par with the one that we inhabit: they all belong to the category of possible worlds, and ours does, too; they are all points in logical space, and ours is, too. Our world, Lewis maintains, we may call 'the actual world', but actuality is nothing more than a relative property: for every individual a in logical space, the actual world is the world that a inhabits.

Lewis contended, moreover, that fictional entities are non-actual individuals.[20] On his view, Ireneo, for instance, is a flesh-and-blood man who possesses an extraordinary memory and extraordinary perceptual abilities; he had a serious equestrian accident, and he suffers from insomnia—but he exists in a region of spacetime that is radically disconnected from the one that we inhabit.[21]

---

20. In his (1986, pp. 91–92), Lewis called non-actual individuals 'alien individuals'.

21. The thesis that fictional entities are non-actual individuals Lewis paired it with his counterpart theory. As Haraldsen (2017, p. 343) notes, however, these two doctrines are not wedded to one another: the thesis that fictional entities are non-actual individuals can be consistently paired with the transworld identity theory too. Nevertheless, the statement that fictional entities are non-actual individuals has a slightly diverging meaning for the transworld identity theorist and the counterpart theorist. For the former, a non-actual individual is an individual that does not exist in the actual world; for the latter, a non-actual individual is an individual who does not have a counterpart in the actual world, nor is constituted by individuals (e.g., particles) with counterparts in the actual world. What exactly a fictional entity is on the possibilist's view depends, therefore, on which theory of transworld identification she adheres to. Below in this section there are presentations of both the counterpart theory and the

Kripke advanced a forceful objection against possibilism in his Princeton lectures in 1970, and again in his John Locke Lectures in 1973. It can now be found in print in his (1980, pp. 156–158) and his (2013, pp. 41–53). In order to reconstruct Kripke's objection, I shall introduce one simple piece of terminology: for every *fictum a*, let *a's fictional properties* be the properties that characterise *a* in the story where *a* figures. Ireneo's fictional properties, for instance, are, amongst others, the following: being a flesh-and-blood man; being Uruguayan; being a hypermnesiac; and suffering from insomnia. Now let I be the set of Ireneo's fictional properties. Kripke observed that there are, across various possible worlds, multiple distinct non-actual individuals that possess I. Despite sharing I, however, these individuals differ from one another in regard to a multitude of qualities: one of them, for instance, was ill with pertussis during infancy, and another one was not; one of them lit a cigarette seven seconds before midnight on June 13th, 1885, and another one did not. Which one of them is Ireneo? For each one of these individuals, there is no reason whatsoever to suppose that Ireneo is identical to him and not any one of the others, for Borges's story is entirely silent as to whether Ireneo possesses any one of the aforementioned qualities or not. It follows that, for every one of the individuals that we are considering, we cannot establish that he, and not any one of the others, is Ireneo. Thus Kripke contends that Ireneo cannot be identified with a non-actual individual at all. If we generalise this contention in the obvious manner to the totality of fictional entities, we will obtain the conclusion that possibilism is untrue: fictional entities are not non-actual individuals.

Lewis (1983, pp. 6–17) replied to Kripke by appealing to his own counterpart theory (1971; 1973; 1986, pp. 192–263). This theory says that all individuals in logical space are world-bound: each one of them exist in only one world. Moreover, this theory encompasses a peculiar account of the semantics of names within modal contexts. Suppose that a is an actual individual who is not G. Since individuals are world-bound, Lewis maintains, the statement 'Possibly a is G' must be interpreted as saying not that there is a possible world where a is G, but that there is a possible world where a *counterpart* of a is G. A counterpart of a, according to Lewis, is a non-actual individual that is not identical to a, but is not merely distinct either: it is similar to a in some special respects, and we may talk about it felicitously as if it were a itself. Lewis recognised that there are infinitely many worlds such that, in each one of them, there is a man who bears I. On his view, these men are all world-bound, and thus distinct from one

transworld identity theory.

another; but they are also *counterparts by acquaintance* of one another: each one of them, that is to say, is the man called 'Ireneo' that his world-mates would learn about if they read the series of sentences that constitute Borges's story. Thus, in a peculiar sense, they can all be said to be Ireneo.

Lewis's counterpart theory is subject to a host of objections (see, e.g., Feldman, 1971; Kripke, 1980, p. 45), however, and often repudiated on the basis of them. Additionally, the counterpart theory is wedded to a descriptivist conception of names, which is rejected to a very large extent nowadays. Since Lewis's defence of possibilism depends on this theory, the defence in question is generally considered unsuccessful.

The alternative to Lewis's counterpart theory is the transworld identity theory, which states that individuals exist at various worlds (see, e.g., Kripke, 1980, pp. 41–54). Let us suppose that the possibilist adopts this theory. Now she believes that there exists, in a world w_1, a man who instantiates I and was ill with pertussis during infancy; and she also believes that this man is numerically identical to an man that inhabits $w_2 \neq w_1$ and instantiates I, but was never ill with pertussis. Borges never determined whether his protagonist was ever ill with pertussis or not, but the possibilist need not worry about that anymore: since the men mentioned above are one and the same, it makes no sense to wonder which one of them is Ireneo. Sadly, the possibilist who has adopted the transworld identity theory is subject to a reiteration of Kripke's objection. Even if individuals exist in various worlds, there are, in logical space, multiple non-actual persons who are distinct from one another and bear I.[22] Thus Kripke's question arises again: which one of these persons is Ireneo? The possibilist, once more, will not be able to answer.

It might appear, therefore, that possibilism has been effectively defeated by Kripke.

---

22. That the non-actual individuals that bear I are not one but many follows from two elementary tenets of the possibilist's conception of logical space. The first is that non-actual individuals exist (see Lewis, 1986, pp. 91–92); the second, that actuality is a relative property. If non-actual individuals exist, then there is a non-actual world w_1 where all the inhabitants of the actual world exist, but there also exists a non-actual individual who bears I. Let it be called 'a'. If actuality is a relative property, then nothing is special about the world that we call 'actual'—in fact, every world is actual relative to its own inhabitants. It follows that, for every world w, there must be individuals that are non-actual relative to w; and, in particular, there must be individuals that are non-actual relative to w_1. Since a exists in w_1, a is not a non-actual individual relative to w_1, so there must exist a world $w_2 \neq w_1$ where an individual bearing I exists who is non-actual relative to w_1. Let it be called 'b'. Clearly, b does not exist in our world. Thus it has been demonstrated that the individuals bearing I that are non-actual relative to our world are multiple.

Nevertheless, Caplan (2016, p. 345) and Haraldsen (2017; 2020) have recently suggested that the possibilist may surmount Kripke's offence by adopting AR.[23] Let us see.

2   Arbitrary Reference

As I said in the previous section, Lewis's counterpart theory yields a functional rejoinder to Kripke's objection. If the theory in question were successful, then a different rejoinder to Kripke would be wholly redundant. Caplan's and Haraldsen's AR-based rejoinder, then, occurs under the assumption that the counterpart theory is, as many believe it to be, unsuccessful. Accordingly, in what follows I will assume the truth of the negation of the counterpart theory: I will assume, that is to say, the truth of the trans-world identity theory.[24]

Breckenridge and Magidor (2012) maintain that the reference of instantial terms is fixed in an arbitrary manner. Consider a reasoner who reaches a conclusion conforming to the following schema: '$\exists x \Phi x$'. She may derive 'Φa' by existential instantiation—provided that 'a' does not occur earlier in her proof nor in the proof's conclusion. Within this deductive process, 'a' functions as an instantial term. Informally, 'a' would be introduced to the proof in question by means of a locution akin to the following: 'Let a be a Φ'. Pre-theoretically, then, we could say that 'a' is any one of the Φs, or an arbitrary Φ—but, in reality, the semantics of 'a' is far from clear: the individuals satisfying 'Φx' may be more than one, and at no point is any one of them specified to be the referent of 'a'. Hence, our reasoner intends 'a' to refer to a Φ, but no Φ in particular. Breckenridge

---

23. Stone (2010) comprises a rejoinder to Kripke that differs significantly from Caplan's and Haraldsen's. He suggests that 'Ireneo' refers indeterminately to all of our Ireneo-candidates, and that the truth-values of fictive sentences containing occurrences of 'Ireneo' ought to be assigned a truth-value by supervaluation. To my mind, however, Stone's rejoinder cannot be satisfactory for the possibilist: supervaluation may well allow us to assign truth-values to fictional sentences; but, if there is no non-actual individual that 'Ireneo' determinately refers to, then the possibilist cannot claim that Ireneo is a non-actual individual—and that is exactly what she wants to do. Haraldsen (2017, pp. 354; 2020, p. 404), too, finds Stone's proposal unconvincing, but I shall not reconstruct his arguments here.

24. Importantly, construing possible worlds in a concretist manner is crucial for the possibilist—even if she adopts the trans-world identity theory. In §3, I will show that identifying fictional entities as concretist non-actual individuals allows the possibilist to attribute to *ficta* their fictional properties. If she identified fictional entities with *ersatz* non-actual individuals instead, then she would not be able to do this: one of Ireneo's fictional properties is that of being hypermnesiac, but abstract entities cannot instantiate the property of being hypermnesiac. The ability to attribute *ficta* their fictional properties, moreover, is rare amongst fictional realists: cf. Chapter Two. (For sensible treatments of the various metaphysical worries that may arise by pairing concretism with the trans-world identity theory, see McDaniel (2004) and Yagisawa (2010)).

and Magidor (2012, p. 377) submit that, in virtue of the reasoner's intentions, and as long as there are Φs, 'a' refers arbitrarily: it refers to a Φ, but nothing determines which Φ it refers to.

The following is Breckenridge's and Magidor's technical formulation of the thesis that the reference of instantial terms is not determined by anything at all:

AR Thesis

Let w_a be a possible world where a reasoner r reaches '$\exists x \Phi x$', and derives 'Φe' by existential instantiation at time t. Suppose, moreover, that the instantial term 'e' refers arbitrarily to an individual s satisfying 'Φx'. Then, if there are several Φs, there is a possible world $w_b \neq w_a$ that is absolutely indistinguishable from w_a, except because, in w_b, the instantial term 'e', introduced by r at time t, refers to $s' \neq s$ (Breckenridge and Magidor, 2012, p. 380).

I will refer back to (AR Thesis) later on. For now, let us see how adopting AR may help the possibilist to surmount Kripke's objection.

Haraldsen (2017; 2020) contends that fictive names behave semantically just like instantial terms. Kripke, we have seen, noted that there are multiple individuals in logical space who possess I. According to Haraldsen, Borges intended the name 'Ireneo' to refer to any one of them, and, in virtue of these intentions, the name in question came to refer arbitrarily to one such individual. Caplan confessed to being only relatively sympathetic to this proposal: he finds the idea somewhat implausible that certain reference facts are not determined by anything whatsoever (2016, p. 345). And yet, he thinks that, if AR is true, then adopting it would allow possibilists to overcome Kripke's objection.

I am very sympathetic to Caplan in his reservations concerning AR: the idea is indeed rather incredible that the reference of some referring expressions is not determined by anything whatsoever. To my mind, though, the burden of Breckenridge's and Magidor's (2012) is precisely that reservations of this kind are outright misplaced.[25] In any case, my aim in this section is not that of determining whether Breckenridge and Magidor are right or not: my present aim is, instead, that of investigating whether or not AR may help possibilists to surmount Kripke's objection. For the sake of this investigation,

---

25. See also Kearns and Magidor (2012) for a defence of the thesis that there are brute semantic facts.

I shall assume, for now, that AR is correct. If it is correct, then the possibilist may claim, with Haraldsen, that Borges intended the name 'Ireneo' to refer to any one of the a non-actual individuals who bear I, and the name did exactly so. Thus Kripke's offence dissolves: the name 'Ireneo' does refer to a particular *possibilium*; there is a particular *possibilium* that Ireneo is identical to.

If the possibilist generalises Haraldsen's proposal in the obvious manner, she will be able to assert, contra Kripke, that every fictive name refers to a non-actual individual and, consequently, that every fictional entity is identical to a non-actual individual. In the following section, however, I will argue that having adopted AR will lead the possibilist astray: it will lead her, in particular, to betray her own doctrine.

3   P+AR

We have seen that, for every *fictum a*, *a*'s fictional properties are the properties that characterise a in the story where a figures. Let us now say that fictional sentences are the sentences where fictional entities are attributed their own fictional properties. (1) and (2), then, are both fictional sentences: both are sentences where Ireneo is attributed his own fictional properties. Let us say, moreover, that creation sentences are sentences where fictional entities are attributed the property of having been created by the authors that created them. Thus (3) is a creation sentence.

Fictional sentences are intuitively true: witness, for instance, (1) and (2). Creation sentences are intuitively true, too: (3) illustrates that this is so. Nevertheless, it is truly challenging to construe, at once, fictional and creation sentences as being both true. Consider, for example, a couple of fictional sentences and a creation one: (1), (2) and (3). If these three sentences are true, then Ireneo possesses sheerly conflicting qualities: he is, at once, a hypermnesiac Uruguayan man and Borges's authorial creation. The following, however, seems irrefutably true: a writer's authorial endeavours simply cannot result in the generation of Uruguayan men; whatever the products are of the authorial endeavours of writers, these products are not Uruguayan men. If these two statements are true indeed, then how could (1), (2) and (3) be all true? Construing fictional and creation sentences as being both true is, in sum, not at all an easy task. In fact, as I showed in Chapter Two, fictional realists have gone to great lengths to try to complete said task successfully and they have failed.

Let P+AR be the sum of possibilism and AR. In the remainder of this section, I shall

pursue an answer to the following question: is the P+ARist capable of successfully carrying out the aforementioned task?

Lewis countenanced a wealth of vastly diverse non-actual individuals. Armed with AR, however, the P+ARist may claim that the name 'Ireneo' refers to one of them exactly: it refers arbitrarily to one of the multiple non-actual individuals who exist in logical space and bear I. The non-actual individual that 'Ireneo' refers to, then, is, among other things, an hypermnesiac Uruguayan man. Consequently, within the P+ARist framework, the fictional sentences (1) and (2) are true, as our intuitions say they are. And, in more general terms: for every *fictum a*, the P+ARist may resort to the theory of arbitrary reference to contend that '*a*' refers arbitrarily to one of the various non-actual individuals who possess *a*'s fictional properties. A trivial result of this contention is the following: within the P+ARist framework, every *fictum a* possesses *a*'s fictional properties. If this is the case, then fictional sentences are true within the P+ARist framework: by definition, a fictional sentence is a sentence where a fictional entity is attributed its own fictional properties; if all fictional entities possess their own fictional properties, then fictional sentences are all true.

What about creation sentences, though? According to Lewis, possible worlds are causally isolated from one another. If he is right, then an actual individual is radically precluded from causing a non-actual individual to exist. If fictional entities are non-actual individuals, as the P+ARist says they are, then they were not brought into existence by authors in the actual world. It follows, it appears, that creation sentences are all untrue—but Haraldsen maintains that this is not the case.

On Haraldsen's view (2017, pp. 346–349), for every *fictum a*, the act that we refer to as the act of creating *a* is, in reality, the act of *selecting a creatively* amongst a wealth of vastly diverse non-actual individuals. Accordingly, creation sentences, on Haraldsen's view, concern not acts of generation, but acts of creative selection. If the the acts in question did indeed take place, as Haraldsen says they did, then creation sentences are all true.

To my mind, the thesis may well be true that some of the acts that we refer to as acts of creation are acts of creative selection.[26] Even if it is, however, the P+ARist cannot

---

26. A thesis very much alike this one is compellingly defended in, e.g., Kivy (1983; 1987); Dodd (2000; 2002; 2007).

hold creation sentences to be true. Consider (3), for example. If the name 'Ireneo' in (3) refers arbitrarily, as the P+ARist says it does, then there is simply no sense in which Borges selected Ireneo amongst other non-actual individuals. On the theory of arbitrary reference, we have seen, arbitrarily referring expressions are those whose reference is not determined by anything whatsoever. Trivially, if the reference of an expression is not determined by anything whatsoever, then it is not the case that somebody selected the reference of the expression in question. Suppose, for the sake of contradiction, that an expression 'a' refers to an individual b, and suppose that b was selected as the reference of 'a' by an individual c. If this is the case, then there exists something that determined what the reference of 'a' is: what the reference of 'a' is was determined by c's will. Hence, if the reference of an expression is not determined by anything whatsoever, then it is not the case that somebody selected the reference of the expression in question.

The conclusion is thus inescapable that Borges did not select the reference of 'Ireneo' amongst other non-actual individuals at all. (3), then, is untrue; and, by analogous arguments, all other creation sentences are untrue as well.

Let us, though, be charitable, and investigate whether the advocate of P+AR can some-how avoid this result. She thinks that the name 'Ireneo' came to refer to Ireneo when Borges wished it to refer arbitrarily to any one of the multiple non-actual individuals who bear I. If this is right, then the individual that 'Ireneo' ultimately referred to was not selected by Borges amongst the other non-actual individuals who bear I: the reference of 'Ireneo' was fixed to one of these individuals, and not the others, not in virtue of Borges's will, but arbitrarily. Nevertheless, there is something that Borges did creatively select: he creatively selected I—the collection of properties, that is to say, that he wished were borne by the referent of 'Ireneo'. If (3) describes this act of creative selection, however, then the referent of 'Ireneo', as it occurs in (3), is not a non-actual individual, but a set of properties.

Platonists think that fictional entities are atemporal *abstracta* (see, e.g., Wolterstorff, 1980). According to Friedell (2020, p. 224), a "representative version" of the Platonist doctrine states that *ficta* are identical to the sets of their fictional properties. The advocate of P+AR could, I believe, turn to Platonism in order to contend that she can, after all, comply with our intuitions that (3) is true. Some occurrences of 'Ireneo', she could tell us, refer to a non-actual individual, but others refer to a certain set of properties. The occurrence of 'Ireneo' in (3) is an example of the latter.

The advocate of P+AR, in sum, could read (3) as saying that Borges creatively selected the set of properties that characterise his protagonist. It is undoubtedly true that he did so, and, consequently, (3) is true under this reading.

It seems, in sum, that the P+ARist is capable of construing both fictional sentences and creation sentences as true after all. This feat, though, comes at a cost. In order to achieve it, the advocate of P+AR must adhere to the thesis that each and every fictive name is ambiguous between a non-actual individual and a set of properties; she must admit, then, that there does not exist one thing called 'Ireneo', but two.

Now, if P+AR comprises the ambiguity thesis that I just stated, then P+AR does not count as a possibilist theory under the standard definition of 'possibilism'—the one, that is to say, appealed to by Lewis, Kripke, and Haraldsen and Caplan themselves. Possibilism, under this standard definition, is the view that *ficta* are non-actual individuals and nothing else, and, accordingly, that fictive names refer exclusively to non-actual individuals. Nevertheless, that is patently not what P+AR now says. P+AR now says that fictive names refer to non-actual individuals only on some of their occurrences: on the others, they refer to sets of properties. According to the P+ARist, then, the things that we identify as *ficta* belong to two distinct ontological categories, and the statement that *ficta* are non-actual individuals and nothing else is, consequently, untrue. Consider the case of Ireneo as an example once more: we have seen that, if the possibilist adopts AR, she will be able to contend that the individual that we speak about when we utter (1) and (2) is a *possibilium*. If she wants to hold (3), too, as true, however, she will have to claim that (3) concerns not a *possibilium*, but a set of properties; she will have to contend that the thing that Borges wrote about in 1942 and we identify as a fictional entity is a set of properties—and that, of course, is not a possibilist thesis at all.

In sum: adopting AR may allow the possibilist to surmount Kripke's offence—it may allow her to contend that, for every fictive name, there exists one particular *possibilium* that this name can be used to refer to. Nevertheless, adopting AR will ultimately lead the possibilist astray: it will force her to endorse a theory of the ontology of *ficta* that is not the possibilist theory.

In this section, I showed that the adoption of AR has unwanted consequences for the possibilist. I believe, nonetheless, that there is a promising insight behind Haraldsen's and Caplan's claim that adopting AR would allow the possibilist to surmount Kripke's offence. In the next section I will exploit the insight in question.

4   The Semantics of Instantial Terms

Suppose that a reasoner reaches an instance of '$\exists x \Phi x$', and derives 'Φa'. What does 'a' refer to? The individuals satisfying 'Φx' might be more than one, and our reasoner would defeat her own inference if she chose one of them as the referent of 'a' (cf. Shapiro, 2012, p. 403). Thus she cannot single out a particular Φ amongst the others in order to designate it as the referent of 'a'. Clearly, this situation resembles that of the possibilist who faces Kripke's objection: this possibilist cannot single out a specific non-actual individual in order to designate it as the referent of a fictive name. It seems rational, then, to suppose that a method for conferring a semantic value to instantial terms could help possibilists fix the reference of fictive names. What Caplan and Haraldsen did was to try out one of these methods, but there are others—and, hopefully, another method will not lead the possibilist astray.

Some theories of the semantics of instantial terms construe these terms as empty, and others, as disguised bound variables. Nevertheless, AR is not the only one that construes instantial terms as genuinely referential. According to Fine (1985a; 1985b), there are, besides the totality of particular objects, *arbitrary objects*, and instantial terms refer to them. Besides particular men, for instance, there are arbitrary men; besides particular numbers, arbitrary numbers. Each one of these objects, Fine claims, is associated with a *value-range*: a set of objects whose shared properties determine which properties are borne by the corresponding arbitrary object. Roughly, the attributes that an arbitrary object bears are, on Fine's view, those that are shared by all of the objects that pertain to its value-range.[27]

Perhaps, in order to defend her theory from Kripke's objection, the possibilist could commit to Fine's theory instead of AR. Let P+AO be the conjunction of possibilism and Fine's theory of arbitrary objects. Ideally, on P+AO, the name 'Ireneo' refers not to any one of the non-actual men who possess I, but to the arbitrary object that has all of these men within its value-range. If this is the case, then the reference of 'Ireneo' is fixed, and Kripke's objection collapses. Sadly, this ideal scenario cannot obtain.

At one point, Fine (1985b, p. 35) claimed that the arbitrary object theorist ought to countenance a plurality of arbitrary objects associated with each existing value-range. As Breckenridge and Magidor (2012, pp. 390–392) note, there is are simple and

---

27. See Fine (1985b, pp. 10–17) for important refinements of this thesis.

compelling reason to suppose the Fine is right. Witness:

$$\exists x \exists y (\Phi x \wedge \Phi y \wedge x \neq y) \therefore \Phi a \wedge \Phi b \wedge a \neq b$$

This move consists of an application of the rule of existential instantiation (see, e.g., Shapiro, 2012, pp. 403–410). The move is legitimate as long as neither 'a' nor 'b' occur earlier in the relevant proof, nor in the proof's conclusion. On the theory of arbitrary objects, though, the legitimacy of this move depends on there being a plurality of arbitrary objects associated with the same value-range. Both 'a' and 'b' are instantial terms. On Fine's view, they both refer to arbitrary objects. a is an arbitrary object that has all of the Φs within its value-range, and so is b. If the move in question is to preserve truth, then a and b must be distinct. Hence, there are multiple arbitrary objects associated with the same value-range.

Unfortunately, if there are multiple arbitrary objects associated with the same value-range, then the advocate of P+AO cannot fix the reference of any fictive name. Earlier I said that, according to P+AO, Ireneo is the arbitrary object that has, in its value-range, all of the non-actual men that bear I. If there are multiple such arbitrary objects then the question arises of which one of them is Ireneo?

Let S be the set of non-actual individuals who bear I. Fine says that the properties borne by an arbitrary object are determined by its value-range. The arbitrary objects whose value-range is S, then, possess all of the same properties: they are indistinguishable. If this is the case, then it is impossible to chose any one of them as the one that Ireneo is identical to: as Black (1952) famously noted, an entity that is indistinguishable from others cannot be designated as the referent of any name. Thus the possibilist's plight reiterates: once again she is before a plurality of individuals, and she cannot tell us which one of them is Ireneo.

The possibilist, in sum, cannot resort to Fine's theory of arbitrary objects to overcome Kripke's offence. If she is to overcome it, she must adopt AR. Nevertheless, I argued in §3 that adopting AR will lead the possibilist to betray possibilism: it will lead her to defend a theory of the ontology of fictional entities that is not the possibilist theory. If my arguments are sound, then the prospects of success are seriously disheartening for the project of vindicating possibilism.

Now, in §3 I showed that construing both fictional and creation sentences as true is truly challenging. I also showed that the P+ARist is capable of doing so: she need

only postulate a certain ambiguity in fictive names. The P+ARist, then, is capable of achieving a remarkable feat. If this is the case, then perhaps adopting AR is a virtuous move for the possibilist after all: it may not lead her to vindicate possibilism, but perhaps it does lead her to articulate a theory of the ontology of fictional entities that is powerful and, thus, worth considering.

Nevertheless, in order to achieve her remarkable feat, the P+ARist had to adhere to AR; and, as it happens, AR is truly controversial. Haze (2016), Child (2019) and Gasparri (2021) have all raised multiple arguments against it, and, in the following section, I will raise an original one myself.[28] If my argument is sound, then the P+ARist stands, with her remarkable feat and all, on seriously unstable grounds.

5   Arbitrary Reference and Weak Discernibility

My aim in this section is to establish that AR is severely dubious. In order to do so, I will show that it is inconsistent with a metaphysical thesis that has been robustly defended and is, nowadays, widely accepted: the thesis that weakly discernible entities are metaphysically possible.

Two entities a and b are weakly discernible just in case (i) they are indistinguishable (i.e., they share all of their intrinsic and relational properties), and (ii) there is an irreflexive and symmetric relation R such that Rab. If (i) fails, a and b are strongly discernible; if, instead, (ii) fails, then a and b are utterly indiscernible (cf. Butterfield and Caulton, 2012, p. 45–50; Assadian, 2019, pp. 2551–2552). There are weighty reasons to believe that weakly discernible entities are metaphysically possible. Let w_1 be a world inhabited exclusively by two iron spheres, a and b, which are almost indistinguishable: they share all of their intrinsic and relational properties, except because one of them is singly scratched. There seems to be no reason to believe that such a possible world does not exist. Lewis's principle of recombination (1986, pp. 86–92) states that, for every rearrangement of actual objects or parts thereof, there is a possible world where such rearrangement exists. Evidently, a certain recombination of actual iron atoms would result in the existence of iron spheres satisfying exactly my descriptions of a and b, so this principle allows for the existence of w_1. Armstrong's combinatorialist principle (1989), on the other hand, says, roughly, that, for any particulars and any universals, there exists a possible world where these particulars instantiate these universals, so

28. See Meléndez Gutiérrez (forthcoming), however, for a refutation of Haze's objection against AR.

long as no two of the latter are determinates of the same determinable. It is clear that neither a nor b is supposed to instantiate several determinates of the same determinable, so Armstrong's principle, too, allows for the existence of w_1. Additionally, there would seem to be no way to derive a contradiction from a maximally detailed description of w_1—not even by appeal to any one of the formulations of the principle of identity of indiscernibles.[29] Thus, we may resolve that w_1 is metaphysically possible. Moreover, the sphere that is scratched in w_1 could have been unscratched, so there is a possible world $w_2 \neq w_1$ where a and b are indistinguishable: they share all of their intrinsic and relational properties without exception. Importantly, though, there is a symmetric and irreflexive relation that holds between a and b in w_2: a is at a non-zero distance from b, and vice versa; but a is not at a non-zero distance from itself, nor is b. It follows that a and b in w_2 are weakly discernible. Hence, weakly discernible entities are metaphysically possible.

Hacking (1975) believes that this conclusion is a *non sequitur*. Let a *symmetrical universe* be a possible world that contains nothing more than two weakly discernible entities.[30] According to Hacking, every world that we conceive of as symmetrical can be accounted for as non-symmetrical: w_2, for instance, can be thought to contain nothing more than two weakly discernible iron spheres—as we just did; but it may well be that, in reality, this world is one whose space is non-Euclidean, and is occupied at all times exclusively by a single iron sphere. Perhaps, that is to say, a and b are numerically identical, and what we thought to be a non-zero distance between them is, in reality, the length of the journey around space from $a = b$ to $a = b$ itself. If Hacking is right, we can never be certain there are possible worlds inhabited by weakly discernible entities.

There is, however, a powerful rejoinder to Hacking's contention due to Adams (1979). Perhaps w_2 may be thought of as comprising a single iron sphere in a tightly curved space; but it is clear that w_1 cannot be conceived of similarly. In w_1, there is a pair of iron spheres that differ from one another only in that one of them is singly scratched.

---

29. See, e.g., Forrest (2020) for an outline of all of the various versions of the principle of identity of indiscernibles.

30. A symmetrical universe is canonically defined as a possible world that can be divided into two non-overlapping regions of spacetime such that: these regions share all of their intrinsic properties, and their relational properties are only symmetrical relations to one another. This definition entails that a symmetrical universe is a universe inhabited by nothing other than two weakly discernible entities. Clearly, the two halves of the world that we are considering are indistinguishable (in the sense specified above); and, since there is a non-zero distance between them, they are related by a symmetric and irreflexive relation.

The same iron sphere cannot be both scratched and unscratched, so the iron spheres in w_1 are undoubtedly two. Since the scratched iron sphere in w_1 could have been unscratched, it seems that there must be a possible world where both spheres are absolutely indistinguishable. Hacking, however, hypothesises that the possible world in question is, instead, one where space is tightly curved and there exists one single iron sphere. Evidently, Hacking's hypothesis amounts to the following: had the scratched sphere in w_1 been unscratched, the geometry of space would have been different. This is severely counterintuitive: surely a scratched iron sphere could have been unscratched without space folding onto itself. More importantly, if the iron spheres a and b are numerically identical in w_2, then, by the necessity of identity, they are numerically identical in w_1—but that is absurd because one of them is scratched and the other one is not.

Forrest (2020), moreover, noted that Hacking's strategy for reinterpreting symmetrical worlds as non-symmetrical is rather implausible when applied to worlds that are slightly more complex than w_2. Suppose, for instance, that we describe w_3 as populated by three iron spheres that resemble each other exactly: a, b and c. They are arranged in a straight line, and b, the one in the middle, is equidistant from the other two. According to Forrest, Hacking's project of construing w_3 as non-symmetrical would consist of two steps. First, he would identify the outer spheres by interpreting the space in w_3 as curved. He would be thereby left with two indistinguishable spheres: $a = c$ and b. If he wanted to deny that such indistinguishable spheres exist, he would have to identify $a = c$ with b—but this, Forrest claims, seems problematic, as b seemed initially strongly discerned from a and c by the property of being equidistant from two distinct spheres.

I thus deem as established that the thesis that weakly discernible entities are metaphysically possible is significantly compelling. In what follows, I will show that AR is false under the assumption that this thesis is true. Then I will defend my argument from two potential objections.

In his (1952), Black famously argued that an individual, if it is weakly discernible from others, cannot be referred to. He devised a possible world inhabited exhaustively by two iron spheres that share all of their intrinsic and relational properties. These spheres are weakly discernible: each one of them is at a non-zero distance from the other, but not from itself, so they are related by a symmetric and irreflexive relation. Black then contended that neither of these spheres could ever be baptised: given that they share all

of their properties, none of them may be singled out by description in order to stipulate it to be the referent of an expression. After Black's paper was published, it appeared uncontroversial that it is impossible to refer to an entity that is weakly discernible from others. The ARist, however, believes that it is not impossible. On her view, we cannot intentionally choose a specific entity, if it is weakly discernible from others, as the referent of an expression, but we can use an expression to refer arbitrarily to one such entity. Suppose that there are two weakly discernible individuals, each one of which is characterised exhaustively by 'Φx'. We may introduce an expression with the intention that it refers to any one of the individuals characterised exhaustively by 'Φx'. According to the ARist, the expression will refer arbitrarily to one of the individuals in question in virtue of our intentions.

In conclusion: ordinary reference does not allow us to refer to an entity if it is weakly discernible from others, but we can refer to it with an arbitrarily referring expression. This conclusion, I will show, produces a result that is seriously pernicious for the ARist.[31]

Let us assume that the ARist is right: weakly discernible entities can be referred to by arbitrarily referring expressions. Let w_1 be a symmetrical world inhabited by two of these entities, a and b, each one of which bears properties $G_1...G_n$. Suppose that an individual in w_2 intends the term 'c' to refer to an entity instantiating $G_1...G_n$ in some symmetrical world. By virtue of these intentions, 'c' refers arbitrarily to a; but, clearly, it could have referred to b instead. There is a possible world w_3, that is to say, where 'c' refers to b. This result is evidently vindicated by (AR Thesis).

To my mind, however, this result is severely problematic. If a is the referent of 'c' in w_2, then a possesses a property that b does not—namely, the property of being the referent of 'c' in w_2. And, conversely, b bears a property that a does not: b is the referent of 'c' in w_3, but a is not. If this is the case, however, then a and b are not weakly discernible, but strongly so. Importantly, w_1 is an arbitrary symmetrical universe, so the strong discernibility result generalises universally: for every symmetrical universe w, the inhabitants of w are strongly discernible rather than weakly so. If this is the case, though, then there is no possible world inhabited by weakly discernible entities; weakly discernible entities, equivalently, are metaphysically impossible. The problem is

---

31. At this point, an objector could protest that my conclusion is unsound. She could contend that, even if AR is untrue and there is no arbitrary reference, it is possible, *pace* Black, to refer to entities that are weakly discernible from others. Her argument goes as follows:

that, as I argued above, the thesis that weakly discernible entities are metaphysically possible is strongly compelling. If it is true, as it appears to be, then the theory of arbitrary reference is false.

I suppose that the arbitrary reference theorist could try to reject this result by contending that, even if a, and not b, is referred to by 'c' in w_2, there is, for some reason or another, no genuine relation between 'c' in w_2 and a, and, consequently, no genuine property borne by a that is not borne by b as well.

Nevertheless, the ARist cannot defend this thesis without incurring inconsistency, for the very formulation of the theory of arbitrary reference seems to assume that semantic properties (e.g., the property of referring to a particular individual, and its inverse,

---

Let w_1 be a symmetric universe populated exclusively by two iron spheres. These spheres are indistinguishable in w_1, but, clearly, there is a world $w_2 \neq w_1$ where they are not: in w_2, one of the spheres is singly scratched. Let us call the scratched sphere 'a', and the other one, 'b'. Under the assumption that the transworld identity theory is true, the spheres in w_2 and the spheres in w_1 are numerically identical. It follows that the spheres in w_1, despite being indistinguishable, bear names now: one of them I just called 'a', and the other one, 'b'. Since w_1 is an arbitrary symmetrical universe, we may validly conclude that each member of every pair of indistinguishable entities can be baptised, so to speak, through a modally indirect method: we need simply consider a world where these entities are discerned by a single property, baptise them, and then go back to the original, symmetrical world. Now we can refer to this world's indistinguishable denizens by their new names.

Unfortunately, this argument is unsound: its conclusion depends on a surreptitious appeal to AR. Let us pay close attention to the point where my objector asserts that a world w_2 exists where one of the spheres in w_1 is singly scratched. Clearly, 'w_2', as it occurs in this assertion, is an instantial term. If AR is not true, then this term does not refer arbitrarily to a possible world. In fact, if AR is untrue, 'w_2' does not refer to a possible world at all: amongst the various theories that exist of the semantics of instantial terms, AR is the only one that says that instantial terms refer to individuals pertaining to the realm that possible worlds pertain to. As Breckenridge and Magidor themselves recognise (2012, pp. 383–393), alternative theories state that instantial terms are semantically empty; that they are disguised bound variables (see, e.g., King, 1991); or that they refer to queer objects called 'arbitrary objects' (see, e.g., Fine, 1985a; 1985b). It is unquestionable, then, by the ARist's own lights, that, if AR is untrue, then 'w_2' does not refer to a possible world at all. Let us suppose that the semantics of the instantial term 'w_2' is, instead, that which King (1991) would say it is: let us suppose that 'w_2' is a disguised bound variable. Now us go back to my objector's claim that there is a possible world w_2 where one of the spheres is singly scratched. The objector wishes 'a' to be the name of the one sphere in w_2 that is scratched, but that desire is deeply confused. If 'w_2' is a disguised bound variable, then the only thing that my objector actually claimed is that one of the spheres possesses the property picked out by the following propositional function:

There exists a possible world w such that x is singly scratched in w.

But, of course, the other sphere possesses this property as well: there is a possible world where the other sphere is the one that is singly scratched. If this is the case, then the property in question does not discern the denizens of w_1: it is both of them that bear it. Thus the following question arises: how could we then single one of these denizens out in order to designate it as the referent of a name? The answer, of course, is that we cannot. I therefore submit that my objector is wrong: if AR is untrue, and there is no arbitrary reference, then, just as Black maintained, it is impossible to refer to entities that are indistinguishable from others.

being referred to by a particular expression) are genuine, discerning properties. (AR Thesis), we have seen, spells out in technical terms what it is for an arbitrary referring expression to have its reference determined by nothing at all. Several instantial terms occur in (AR Thesis). According to the ARist, these terms refer arbitrarily: 'w_a', for instance, refers arbitrarily to a possible world where, at time t, a reasoner r uses an instantial term in order to refer arbitrarily to one of the Φs. The crux of (AR Thesis) is that there is a possible world $w_b \neq w_a$ that differs qualitatively from w_a only in virtue of one semantic property: the instantial term used by reasoner r at time t in w_a refers to s, whilst, in w_b, it refers to $s' \neq s$. The ARist, then, intends the instantial term 'w_b' in (AR Thesis) to refer to a possible world of a very specific kind: it must be distinguishable from w_a exclusively in virtue of one particular semantic property. The ARist, therefore, ought to appeal to semantic properties in order to define the class of things amongst which 'w_b' is to arbitrarily find a referent. But then semantic properties must be discerning: they must be such that they allow us to determinately partition a domain into the set of individuals that bear them, and the set of individuals that do not. It follows that we speak truly when we assert that certain entities bear certain semantic properties, and that other entities do not. If this is the case, then it is entirely correct to claim that the semantic property of being the referent of 'c' in w_2 is borne by a and not by b and, thus, that a and b are strongly discernible. The ARist, in sum, is forced, by their own commitment with (AR Thesis), to accept my conclusion that, under the assumption that AR is true, weakly discernible entities are metaphysically impossible. Equivalently: AR is inconsistent with the thesis that weakly discernible entities are metaphysically possible.

Our intuitions may oppose the result that a and b are strongly discerned by the semantic relation of being the referent of 'c' in w_2; our intuitions might (on occasion, at least) appear to suggest that a and b, two entities that share all of their non-semantic intrinsic and relational properties and, hence, cannot be told apart, are indistinguishable even if they possess different semantic properties. What I have just demonstrated, however, is that it is not the inconsistency result that I have proved that originally betrays these intuitions, but the theory of arbitrary reference itself: it is this theory itself, and (AR Thesis) in particular, that entails that semantic properties are discerning. As a consequence, the intuitions in question cannot be adduced to defend AR.

In this section, I argued that AR is inconsistent with the thesis that weakly discernible entities are metaphysically possible. This thesis is strongly compelling and, conse-

quently, AR is severely dubious.

In §3, I showed that the P+ARist is capable of achieving a remarkable feat: she is capable of construing fictional and creation sentences as being both true. This achievement, though, depends on her endorsement of AR. If the arguments that I advanced in this section are sound, then the P+ARist, with her remarkable feat and all, stands on seriously unstable grounds.

6   Conclusion

In this chapter, I argued that possibilism cannot be vindicated on the face of Kripke's objection by way of an appeal to AR. For one thing, adopting AR will lead the possibilist to betray her own doctrine. For another, AR is severely dubious and, hence, the possibilist who adheres to AR stands on unstable grounds. In §4, moreover, I argued that possibilism cannot be vindicated either by appealing to Fine's theory of arbitrary objects. If my arguments are sound, then the prospects of success for the project of vindicating possibilism are much more disheartening than they appeared.

In Chapter Two, I contended, that creationism is not satisfactory. If, additionally, the project of vindicating possibilism is disheartening, then, I believe, it is worth investigating whether a different theory can fare better in the challenge of accounting for the ontology of *ficta*. Accordingly, the following three chapters I will devote to formulating, defending and motivating an original theory thereof: the type theory.

Chapter Four:
Types

In Chapter Five, I will advance an original theory of the ontology of *ficta*: the type theory. The type theory states that *ficta* are types of concretist *possibilia*. With the purpose of introducing this theory in the chapter that follows, I will devote the present chapter to the task of defining types and defending their existence. In §1, I will elucidate the ontological nature of types, and I will present their existence and identity conditions. In §2, I will discuss the view that types do not exist, and I will argue that it is untenable. Finally, in §3, I will develop the realist theory of types, and I will defend it from its more prominent objections.

1   The Ontological Nature of Types

In the most minimal of their characterisations, types are abstract entities capable of having instances. The instances of a type are called 'tokens'. Wetzel (2018) formulated what seem to me to be two admirably limpid illustrations of what types are. If a mother tells a paediatrician that her infant has uttered three hundred words, the paediatrician will wonder whether these words are word-tokens or word-types, for the truth of the latter disjunct would indicate that the infant in question is a prodigy. And similarly: consider the title of Timothy Pratt's 'From the Andes to Epcot: the Adventures of an 8,000-Year-Old Bean' (1998). The question naturally arises whether the bean at issue is a bean-type or a bean-token.

Let us say that those who believe in the existence of types are type realists—and henceforth simply 'realists'. Realists infer the existence of types form the existence of sentences of a particular sort. Consider, for instance:

    1.   The CDK4 protein inhibits the p14 protein.

2.   'A' is a letter of Phoenician origin.

(1) and (2) are intuitively true—but what are they about? Every particular, spatiotemporally located biomolecule pertaining to the class of the CDK4 proteins is a CDK4-token. Now, the expression 'the CDK4 protein' in (1) is a singular designator—it designates, that is to say, one single individual. Nevertheless, it is unquestionable that the individual that the expression in question designates is not a CDK4-token: clearly, there is no particular, spatiotemporally located biomolecule that (1) is about. It follows, the realist contends, that the expression at issue refers to the CDK4-type.

Now, entities of various kinds are 'A'-tokens: carvings on stone, wood and bone; ink marks, graphite marks, crayon marks, and aggregates of flared pixels on electronic screens. The expression "'A'" in (2) is a singular designator, but it is clear that it does not refer to any one of the aforementioned entities: (2) does not say, of any one of these entities, that it is of Phoenician origin. Hence, the realist concludes, the expression in question refers to the 'A'-type.

In sum, therefore, the realist's primary argument for the existence of types is the following: there are intuitively true sentences each of which seems to involve reference to an individual of a certain sort—and yet, amongst the tokens of the sort in question, there is not one which the individual in question could be identified with. Consequently, we ought to identify this individual with a type and, thus, admit types into our ontology.

But what, exactly, is a type? Earlier I said that types are instantiatable *abstracta*, but they cannot be exhaustively defined as such, for properties, too, are abstract and instantiatable.[32] One way of defining types more precisely consists in identifying each type with the equivalence class of its tokens. Consider, for instance, the eight-thousand-year-old bean referred to in the title of Pratt's aforementioned article. As it turns out, the bean in question is a type of bean, and, in particular, the one tokened by nuña beans. Now consider the set of all living beings that exist in the actual world. There is a relation R that partitions this set into equivalence classes each one of which contains all and only those individuals that pertain to one and the same species. According to the equivalence-class conception of types (henceforth 'EC'), the nuña-type is one of these equivalence classes: the one, in particular, which contains all and only nuña-tokens. Let

_____

32. At least on the standard understanding of their ontological status—cf., e.g., Orila & Paolini Paoletti (2022, §2.4). See Orila & Paolini Paoletti (2022) also for an overview of alternative views.

this class be called 'N'.

The identification of types with equivalence classes fosters a clear understanding of the ontological nature of types, but it faces a number of difficulties. Equivalence classes are sets, and sets have extensional identity conditions, but it has been noted repeatedly that types do not (see, e.g., Wollheim, 1980, p. 50; Wolterstorff, 1980, pp. 44–45; Dodd, 2007, pp. 40–41; Wetzel, 2018, §4.1.1). Let w_1 be the possible world where one of the nuña-tokens that actually exists fails to exist. By the axiom of extensionality, the equivalence class of the nuña-tokens in w_1 is not N but $N' \neq N$. If the nuña-type is identical to the equivalence class of nuña-tokens, then the nuña-type in the actual world is distinct from the nuña-type in w_1. Sadly, this result is unacceptable: surely a nuña-token could have failed to exist without Pratt writing his article on a wholly different matter.

The advocate of EC, though, may easily adjust her view to so that it complies with the non-extensionailty of types: instead of commencing her construal of types with the set of all actual living beings, she must do so with the set of all *possible* living beings.[33] Even if she did so, however, the identification of types with equivalence classes will remain subject to a significant complaint. When the hypothetical paediatrician that I mentioned above wonders whether her patient has uttered three hundred word-types or word-tokens, she is wondering whether the infant's vocabulary comprises three hundred elements, or whether it comprises fewer elements, but of some of them the infant has produced multiple *occurrences* or *embodiments*. This fact illustrates the statement that I advanced at the beginning of this chapter: a defining characteristic of types is that of being instantiatable. Nevertheless, the status of tokens as instances of types appears rather neglected on EC, for, on EC, the relation that relates a type with its tokens is not instantiation but membership.

Perhaps EC could be amended somehow in order to honour the fact that types are related to their tokens by the relation of instantiation. As long as such amendment is not formulated, however, a reasonable course of action for us to take consists in identifying types not with equivalence classes, but with *sui generis universals*. This course of action has been followed by, e.g., Wollheim (1980, p. 50); Wolterstorff (1980,

---

33. If types are non-extensional, as I have shown that they are, they must be intensional or hyperintensional. I shall not take a stand on this issue, however, for settling it requires an extensive discussion that is irrelevant for my present purposes.

pp. 44–58); Dodd (2007, pp. 11, 37–48), Wetzel (2009, pp. xi-xii), and, in fact, by most metaphysicians (cf. Wetzel, 2018, §3). It is also the course of action that I will henceforth follow.[34]

On the conception of types that I endorse, then, types are *universals:* they are, that is to say, abstract entities characterised by the property of being instantiatable. They are, moreover, *sui-generis:* they constitute a kind of universal that is determinately distinct from other kinds thereof—and distinct, in particular, from properties.

According to Wollheim (1980, pp. 50–51), types are different from properties insofar as types, but not properties, possess two special qualities—namely, the qualities that an individual possesses iff it satisfies both of the following propositional functions:

    i. x shares properties with its instances.

    ii. x is such that we systematically think and speak of it as if it were an archetypical instance of x.

Let us use a couple of examples to illustrate Wollheim's claim that types, but not properties, satisfy (*i*) and (*ii*). Consider:

    3.   The property of being a tiger is a large felid that bears an orange, striped coat.

(3) is clearly untrue: the property of being a tiger is not a large felid, and it does not

---

34. As I said earlier, I am introducing types in this chapter with the purpose of identifying *ficta* with types in the next one. For this purpose, it is ultimately irrelevant whether types are equivalence classes or *sui generis* universals. If the equivalence-class conception of types could be wholly satisfactorily formulated, then the types that I will identify with *ficta* could be unproblematically conceived of as equivalence classes. Given the availability of an alternative conception, however, I shall not attempt to produce the formulation in question.

Importantly, though, there is a rather daunting question that advocate of EC ought to respond for EC to be considered satisfactory: what are the relations under which types, *qua* equivalence classes, are defined? With regard to some types, the answer is easy to find. For instance: the defender of the equivalence-class conception believes that the type of the perfectly vertical lines is the equivalence class that comprises all possible perfectly vertical line-tokens. This equivalence class is obtained by partitioning the set of all possible lines under the relation of parallelism. With regard to other types, however, finding an answer to the aforestated question is not easy at all. Consider, for example, the Oxford-shoe-type. According to EC, this type is the equivalence class that comprises all possible Oxford-shoe-tokens, and this equivalence class is obtained by partitioning the set of all possible shoes under a relation R—but what relation is R? Not, of course, the relation that x bears to y iff x and y are equally coloured shoes; not the relation that x bears to y iff x and y are equally sized shoes. The answer, it appears, is this: R is the relation that x bears to y iff x and y are of the same *type*. This answer, though, is circular and, thus, useless.

bear an orange, striped coat. The property of being a tiger, then, does not share with its instances any one of the properties that are attributed to it in (3). In fact, it seems truly difficult to think of a property possessed by both the property of being a tiger and its instances—at least if we do not count existence as a property. The property of being a tiger, then, does not satisfy (i)—but it does not satisfy (ii) either. (3) describes the property of being a tiger as if it were an archetypical tiger: it describes it as a large felid that bears an orange, striped coat. (3), however, is intuitively untrue, so we would not assert it. It follows that the property of being a tiger does not satisfy (ii): we do not speak of it as if it were an archetypical tiger.

The tiger-type, by contrast, satisfies both (i) and (ii). Witness:

4.   The tiger is a large felid that bears an orange, striped coat.

Every particular, spatiotemporally located organism pertaining to the species of the *Panthera tigris* is a tiger-token. The locution 'the tiger' in (4), however, does not denote any one of these organisms: there is no particular, spatiotemporally located member of the *Panthera tigris* species that constitutes the subject matter of (4). The locution in question, therefore, denotes not a tiger-token, but the tiger-type. (4), moreover, is intuitively true. If it is true indeed, then the tiger-type, like tiger-tokens, is a large felid that bears an orange, striped coat. The tiger-type and its tokens, then, share properties.

The tiger-type thus satisfies (i)—but it also satisfies (ii). (4), we have seen, describes not a particular tiger-token, but the tiger-type. Nevertheless, the mental representations that the description in question is meant to elicit, it seems, are representations of an archetypical tiger-token: a large animal whose bodily structure is that which characterises felids; a large animal bearing an orange, striped coat. Hence, we think of the tiger-type as an archetypical tiger-token—and we speak of it as if it were an archetypical tiger-token, too: as proven by our impressions that (4) is true, we gladly attribute to the tiger-type the properties that we conceive of as characteristic of archetypical tiger-tokens. The tiger-type, in sum, satisfies both (i) and (ii).

We can now ascertain that types and properties are as Wollheim described them: types satisfy (i) and (ii); properties do not.[35]

35. As a matter of fact, some properties do share properties with their instances. The property of having numerous instances, for instance, has numerous instances, so this property shares with its instances the property of having numerous instances. According to Wollheim, however, the way in which the property

At this point we are in a position to introduce a piece of terminology that will be useful in what follows:

Exemplification

A universal U is such that the instances of U exemplify $U \leftrightarrow U$ satisfies both (i) and (ii).

If U does not satisfy both (i) and (ii), then its instances instantiate it, but they do not exemplify it. The notion of exemplification thus allows us to define types very succinctly as follows: types are universal *abstracta* exemplified by their instances. The notion in question, then, allows us to distinguish types determinately from properties: whilst types are exemplified by their instances, properties are not: properties are merely instantiated by them. And, finally: now that we have determinately distinguished types from properties, we can perspicuously formulate existence and identity conditions for types in terms of properties. For every property G instantiated in at least one possible world, there is a type whose tokens are all and only the individuals in logical space who instantiate G. Moreover, types K and K' are identical just in case the properties that they are associated with are identical.[36]

Two questions arise, of course, at this point: which are the properties that are instantiated in at least one possible world? What are the identity conditions of properties? I am introducing types with the purpose of identifying *ficta* with types in the next

---

in question shares properties with its instances is importantly different from the way in which types shares properties with their tokens. The property of having numerous instances and its instances share the property of having numerous instances *accidentally,* so to speak—simply because it happens to be the case that numerous properties have numerous instances. By contrast, the tiger-type and its tokens do not share the properties that are mentioned in (4) for merely accidental reasons: the fact that the tiger-type possesses these properties is intimately connected to the fact that tiger-tokens possess them. It appears, roughly, that the tiger-type possesses these properties precisely in virtue of the fact that it is the type of the individuals that possess them characteristically.

Importantly, moreover, its seems that some types share some properties accidentally with their tokens. The type of numerously tokened types, for instance, shares with its tokens the property of being numerously tokened, and it does so accidentally—simply because it happens to be the case that numerous types are numerously tokened. In what follows, however, and for the sake of simplicity, I will ignore these rare cases, and use the term 'sharing' to refer only to the non-accidental sort of property sharing. It will soon become clear that the phenomenon of property sharing between types and tokens is somewhat metaphysically enigmatic—but it is so only when the properties in question are shared non-accidentally. Thus all of the discussions that occur in what follows will concern non-accidental property sharing exclusively.

36. These existence and identity conditions for types were first proposed by Dodd (2007, pp. 38–42, 59–69).

chapter. All of the answers to the aforestated questions which have been defended in the relevant literature are consistent with my purpose. I shall, accordingly, refrain from devoting space to privileging any one of these answers over its rivals.

2   Nominalism

The expressions 'the CDK4 protein', "'A'", and 'the tiger' in (1), (2) and (4) have something in common: each one of them appears to refer to an individual of a certain kind, but it does not refer to any one of the tokens of that kind. Each one of these expressions, therefore, appears to refer to a type. Let us say that expressions like these are t-expressions. Let us also say that an intuitively true sentence that comprises a t-expression is a t-sentence. T-sentences, in sum, are those intuitively true sentences which seem to involve reference to types. The realist believes that t-sentences do not only *seem* to involve reference to types but do indeed do so—recall her account of (1), (2) and (4). Nominalists (see, e.g., Goodman & Quine, 1947; Sellars, 1963; Goodman, 1977) disagree. Nominalists deny the existence of abstract entities and, therefore, the existence of types. They contend, accordingly, that t-sentences are not to be read at face value: t-sentences, they maintain, only *appear* to involve reference to types—in reality, they involve only disguised generic quantifications over tokens of the relevant sort.[37] Consider, for instance:

5.   The seriema is a large bird that inhabits the South American forests.

The expression 'the seriema' in (5) does not refer to any one of the seriema-tokens that exist. The realist believes that it refers to a type—but the nominalist disagrees: she thinks that (5) is a disguised generic statement concerning a plurality of seriema-tokens:

6.   Seriemas are large birds that inhabit the South American forests.

The nominalist maintains that her theory is superior to the realist's account of the intuitive truth of t-sentences. Let us consider (5), for example. The realist believes that (5) involves reference to a type. In order to account for the intuitive truth of (5), therefore, the realist postulates a *sui generis abstractum*. The nominalist, by contrast,

---

37. Early nominalists analyses (see, e.g., Sellars, 1963; Goodman, 1977, pp. 258–276) suggested reading these statements not as disguised generics, but as disguised universal generalisations ranging over tokens. That suggestion is easily proved ineffective, however, and surpassed by the generic alternative: see, e.g., Wetzel, (2009, pp. 55–71).

does not: on her view, (5) is to be analysed as (6), and (6) concerns nothing more than a plurality of spatiotemporally located seriemas. Since the nominalist analysis of (5) is ontologically more parsimonious than the realist's, the former is to be preferred over the latter—or so the nominalist says.

In the remainder of this section, I will argue that the nominalist is wrong: her theory is not superior to realism. In fact, I will contend that nominalism is rationally untenable. In order to do so, I will draw from Wetzel's work (2009, pp. 53–103), but a different set of arguments against nominalism can be found in Dodd's (2007, pp. 38–42). Let S be the set of all t-sentences, and let (7) be the schematic t-sentence:

7.   The K is G.

The nominalist proposal is the following:

Generic Nominalistic Analysis
(7) is to be analysed as follows: 'Ks are G'. And, in regimented terms:

$$Gen \ [x \ is \ K] \ [x \ is \ G].^{38}$$

It is easy to demonstrate, however, that not all of the members of S can be analysed in accordance with (Generic Nominalistic Analysis). Consider, for instance:

8.   The phone [a] occurs four times in every utterance of the name 'Axayácatl'.

If (8) is to be analysed in accordance with (Generic Nominalistic Analysis), then it is to be analysed as follows:

9.   [a]-tokens occur four times in every utterance of the name 'Axayácatl'.

And, in regimented terms:

10. $Gen \ x \ [x \ is \ an \ [a]\text{-token}] \ [x$ occurs four times in every utterance of the name 'Axayácatl'].

Sadly, (8) cannot be analysed as (10) because (8) is intuitively true, whilst (10) is false. Each [a]-token is a vocal sound produced by a certain speaker, and it exists at a certain point in spacetime. Similarly sounding [a]-tokens may be easily produced,

---

38. This regimentation coincides with Lewis's (1975) proposal for regimenting generics. See Leslie & Lerner (2022, §1.5) for a summary of this proposal.

but *it* cannot occur again. If this is right, then (10) is untrue: if [a]-tokens do not occur multiply at all, then they do not occur multiply in every utterance of 'Axayácatl'. Suppose, for instance, that I produce an utterance of 'Axayácatl' right now. On January 1st, 1996, Rafael Sebastián Guillén Vicente orally delivered the Fourth Declaration of the Lacandon Jungle. In doing so, he produced a very large number of [a]-tokens, but absolutely none of them occurred in my utterance of 'Axayácatl'—not four times, nor even once. And similarly: in 1953, Juan Rulfo read of his short story *Luvina* aloud in order to create a recording. In doing so, he produced a very large number of [a]-tokens, but not a single one of them occurs in my utterance of 'Axayácatl'—not once, nor, of course, four times. In fact, if [a]-tokens, in some sort of generality, occurred four times in every utterance of 'Axayácatl', then every utterance of 'Axayácatl' would comprise an enormous abundance of [a]-tokens—but that is not how utterances of 'Axayácatl' are. Thus [a]-tokens do not occur, in any sort of generality, in every utterance of 'Axayácatl'. (10), in sum, is untrue—and, hence, not correct as an analysis of (8).

If (8) is to be analysed nominalistically—i.e., as concerning not the [a]-type but [a]-tokens exclusively—, it is to be analysed as follows:

11.   There are four [a]-tokens in every utterance of the name 'Axayácatl'.

(8), then, can be given a nominalistic analysis—but not by way of (Generic Nominalistic Analysis). Importantly, moreover, the members of S which escape (Generic Nominalistic Analysis) are multiple. Witness:

12.   Old Glory had twenty-eight stars in 1846, and now it has fifty.

It is evident that 'Old Glory' in (12) does not refer to any particular, spatiotemporally located flag. The type theorist believes that it refers to a type. Nominalists, however, reject this conclusion, and claim that (12) must be analysed as concerning token-flags only. According to (Generic Nominalistic Analysis), (12) is to be analysed as (13):

13.   Old Glory flags had twenty-eight stars in 1846, and now they have fifty.

Unfortunately, (12) cannot be analysed as (13): (12) is intuitively true, but (13) is not: clearly, it is not the case that twenty-eight-starred Old Glory flags, in any level of generality, underwent a transformation whereby the number of stars figuring in them increased to fifty. It is clear, then, that (12) cannot be analysed in accordance with (Generic Nominalistic Analysis).

A different candidate nominalistic analysis of (12) is the following:

14.   Most Old Glory flags in 1846 had twenty-eight stars; most Old Glory flags now have fifty stars.

Sadly, (12) cannot be analysed as (14). Suppose that a mistake has been made recently in a flag factory, and an awfully large number of Old Glory flags have been manufactured with forty-five stars instead of fifty. As a consequence, most of the Old Glory flags that exist at present have not fifty stars, but forty-five. In this scenario, (14) is patently false, but (12) is not. Consequently, (12) cannot be analysed as (14).

According to Wetzel (2009, p. 86), in order to formulate a satisfactory nominalistic analysis of (12) we must consider two pieces of information. First: a law passed by the United States Congress in 1818 establishes that the number of stars figuring in Old Glory must match the number of states admitted into the Union. And second: the number of States admitted into the Union in 1846 was twenty-eight; the number of States admitted into the Union at present is fifty. Thus Wetzel (2009, p. 86) contends that, if (12) is to be analysed nominalistically, it ought to be analysed along the following lines:

15.   Old Glory flags which conformed with the law in 1846 had twenty-eight stars; Old Glory flags which conform with the law now have fifty stars.

(15) may be regarded as an adequate analysis of (12), and it may be thought to be wholly nominalistic.[39]   Nevertheless, it should be clear at this point that (Generic Nominalistic Analysis) is not applicable to several of the members of S. And it should be clear, moreover, that the members of S for which (Generic Nominalistic Analysis) is not applicable cannot be handled uniformly. Both (8) and (12), for instance, escape (Generic Nominalistic Analysis), and both were ultimately given adequate nominalistic analyses—but these analyses were not reached by parallel methods, nor do they parallel each other at all. In order to reach our analysis of (12), for instance, we had to take legal and geopolitical facts into consideration; not so in order to reach our analysis of (8). Our analysis of (12) concerns those tokens which conform with the law—it concerns, that is to say, law-conforming Old Glory flags. Our analysis of (8), by contrast, does not, for it does not say anything about law-confirming [a]-tokens.

---

39. To my mind, it is severely dubious that 'the law' in (15) refers to a nominalistically acceptable entity; but let us assume, for the sake of the nominalist's contentions, that it does.

It appears, therefore, that there is no systematic method for producing adequate and nominalistic analyses of the totality of the members of S. In other words: it seems that nominalistic analyses may be produced for the members of S only by way of a piecemeal approach: by confectioning, that is to say, nominalistic analyses of the individual members of S in an *ad hoc* manner, one by one.

Clearly, the members of S have not been all nominalised one by one by a piecemeal approach: there are, there is to say, thousands of intuitively true sentences that comprise t-expressions and have not been nominalised by anyone ever—not by me, not by Wetzel, not by Quine, Sellars or Goodman.[40] Now, if a systematic method existed for nominalising the members of S, then we could be certain that they are all nominalisable—even if they had not been all effectively nominalised yet. We could be certain, that is to say, that each $s \in S$ can be nominalised by applying to it the method in question. I have proved, however, that no such method exists. There is, therefore, no certainty whatsoever that the totality of the members of S are nominalisable. In other words: the nominalist's conviction that all of the members of S are nominalisable is nothing more than an article of faith—nothing more, that is to say, than a doctrine that is upheld despite the fact that there is no evidence at all to support it.[41]

By definition, the members of S are intuitively true. Equivalently: we are committed to the truth of the members of S. Some of the members of S, we have seen, have not been nominalised, and we cannot be certain that they are nominalisable. Consequently, there exists at least one $s \in S$ such that: we are committed to the truth of s; under a face-value reading, s is about types; and we are uncertain whether s can be nominalised. In this scenario, we are, as Quine (1961, p. 105) suggests, rationally compelled to admit that we are committed to types. If s refers, under a face-value reading, to types, and we have not proved that s is really about entities other than types, then, as a matter of fact, we are committed to the truth of a sentence that possesses, in effect, no interpretation under which it is not about types. Thus we are, as Quine would contend, committed to types. Of course, we can now work towards an *ad hoc* nominalisation of s, and we might find it. But, as long as the members of S are not all of them nominalised, there will always exist at least an $s' \neq s \in S$ such that: we are committed to the truth of s';

40. Wetzel's (2009, pp. 23–50) comprises a wealth of evidence that the intuitively true sentences that contain t-expressions are indeed thousands.

41. This argument was originally formulated by Wetzel (2009, pp. 54, 82-83).

under a face-value reading s' refers to types; and we have not proved that s' is really not about types. Since we have not demonstrated that s' is not about types, then our commitment to the truth of s' rationally compels us to admit that there are types in our ontology.

At this point, however, the nominalist could protest that the nominalistic analyses of some sentences are truly convincing—even if they do not align with (Generic Nominalistic Analysis)—and, in fact, much more so than the realist's analyses of those sentences. Consider, for instance:

16.   My friend César wants to buy the ODK12 telescope.

Each particular, spatiotemporally located artefact pertaining to the class of the Orion ODK12 telescopes is an ODK12-token, but, evidently, the expression 'the Orion ODK12 telescope' in (16) does not refer to any one of these artefacts: (16) is true despite the fact that there is not one particular ODK12-token that my friend César wants to buy. It follows that the expression in question is a t-expression, and (16), therefore, a t-sentence. According to (Generic Nominalistic Analysis), then, (16) is to be analysed as follows:

17.   My friend César wants to buy ODK12 telescopes.

(17), however, is unacceptable as an analysis of (16). (16) is intuitively true, but (17) is false: my friend César wants to buy an ODK12 telescope, not ODK12 telescopes in any sort of generality. Consequently, (17) cannot be offered as an analysis of (16) at all. If (16) is to be satisfactorily analysed in nominalistic terms, it ought to be analysed as (18):

18.   My friend César wants it to be the case that there exists an x such that x is an ODK12 telescope and César buys x.

(18), the nominalist could contend, is wholly nominalistic, and, importantly, it is much more satisfactory as an analysis of (16) than the realist's analysis. The realist believes that 'the ODK12 telescope' in (16) refers not to any ODK12-token, but to the ODK12-type. She believes, therefore, that (16) states that César wants to buy an abstract object—and that, the nominalist could claim, is untrue. Consequently, the nominalist could conclude, the realist's analysis of (16) is untenable.

My response to the nominalist is the following: I am wholeheartedly convinced that (18) is extremely satisfactory as an analysis of (16). It seems to me that the only complaint that can be raised against (18) is that it entails that César's object of desire is not a telescope, as our intuitions say it is, but the obtaining of a certain states of affairs. To my mind, however, this complaint is not very compelling. There are two things that I must say, however, on behalf of the realist. The first is that the realist's analysis of (16) is not as ludicrous as it might seem to be. It is true, of course, that César does not want to buy an abstract object. In §1, however, I noted that we systematically think and speak of types as if they were archetypical tokens. Perhaps what I say when I assert (16) is that César wants to buy an archetypical ODK12-token—one that is not broken, etc. Since there is no particular archetypical ODK12-token that César wants to buy, I express the idea that he wants an archetypical ODK12-token by using the OKD12-type as a dummy archetypical ODK12.

My primary response to the nominalist, however, is a different one. Perhaps (16) is best analysed as (18). Perhaps, that is to say, (16) is best analysed nominalistically. Even if this is so, though, my argument against nominalism stands: the nominalist has no systematic nominalisation method and, consequently, the thesis that the members of S can be all nominalised is nothing more than an article of faith. In other words: the fact that some t-sentences are best analysed nominalistically does not signify that we can rationally affirm the existence of a nominalistic paraphrase for all of those sentences. If we cannot rationally affirm the existence of said paraphrases, then, as I argued earlier, we are compelled by reason to admit that there are types in our ontology.

I have argued for two interrelated theses in this section. The first states that the nominalist's contention that all of the members of S can be nominalised is nothing more than an article of faith. The second, which follows from the first, states that we are compelled by reason to admit that there are types in our ontology. Articles of faith, of course, cannot be adhered to with rationality, so reason compels us to deny the nominalist's tenet that all of the members of S are nominalisable. Moreover, if reason compels us to admit that there are types in our ontology, then reason compels us to reject the principal nominalist tenet: the tenet, that is to say, that types do not exist. It follows, therefore, that, if my arguments in this section are sound, then nominalism is rationally untenable.

3   Type Realism

The alternative to nominalism is realism: the view, that is to say, that there are types, and that the t-expressions that occur in t-sentences refer to types. If my arguments in §2 are sound, then nominalism is untenable and, hence, realism ought to be endorsed.

At this point, however, the nominalist could raise a challenge for the realist. In the previous section, I showed that the nominalist can formulate satisfactory nominalistic analyses of (5), (8) and (12). But can the realist formulate satisfactory analyses of these sentences? The nominalist could contend that it is dubious that she can. Let us see.

The realist, we have seen, believes that 'the seriema' in (5) refers to the seriema-type. If (5) is true, as our intuitions say it is, then the seriema-type shares with its tokens the property of being a large bird, and the property of inhabiting the South American forests. Nevertheless, the seriema-type is an abstract entity. Abstract entities are not spatiotemporally located and, hence, they seem unfit to instantiate the properties in question: how could an abstract object be large? How could an abstract object inhabit the South American forests? If the seriema-type is indeed unfit to instantiate these properties, then the realist's analysis of (5) is not satisfactory: our intuitions say that it is true, but it is false on the realist's analysis.

On Wetzel's view, the seriema-type is not at all unfit to instantiate the properties in question. In §1, we saw that types share properties with their tokens—recall (*i*). According to Wetzel, the properties that a type K shares with K-tokens, are instantiated by K in virtue of facts concerning K-tokens (2009, pp. 11–12). The seriema-type and its tokens, for instance, share the properties mentioned in (5), and the seriema-type instantiates said properties in virtue of facts concerning seriema-tokens. More specifically: the seriema-type instantiates these properties in virtue of the fact that its tokens instantiate them.

I think that Wetzel's contention is not satisfactory, however. To my mind, the thought that an abstract object instantiates the properties in question is wholly unintelligible. Suppose that an *abstractum a* instantiates the property of being a large bird. What does it mean for a to instantiate this property if a is abstract and, thus, cannot possess a bird's body, cannot fly, and cannot chirp? In what sense is a a bird at all? And how is a large if it is bodyless? Now suppose that a instantiates the property of inhabiting the South American forests. What does it mean for a to instantiate this property if a

is abstract and, therefore, it has never been located in South America? In what sense does a inhabit the South American forests if it is not located in South America at all?

As long as these questions remain unanswered, I contend, we cannot accept Wetzel's thesis that the seriema-type instantiates the properties that are attributed to it in (5). In Chapter Five and, more specifically, in footnote 55, I will explain that the truth of this contention is ultimately irrelevant for my purpose of identifying *ficta* with types. For now, though, let us suppose that the seriema-type does indeed fail to instantiate the relevant properties. If it does, then the following question arises: how is (5) true?

Even if the seriema-type fails to instantiate the properties in question, it is clear that it is related to these properties by a relation that does not relate it to other properties. Intuitively, we speak truthfully when we say that the seriema-type is a large bird that inhabits the South American forests. We do not speak truthfully, by contrast, when we say that the seriema-type is an Arctic sea bird. Witness:

19.   The seriema is an Arctic sea bird.

Consequently, the seriema-type is related to the properties mentioned in (5) by a relation that does not relate it to other qualities—such as the quality of being an Arctic sea bird. And, more specifically: the seriema-type is related to the properties mentioned in (5) by a relation that allows us to speak truthfully when we attribute these properties to it. Let us call this relation 'kinstantiation'. Now consider the following proposal:

Kinstantiation Doctrine
For every type K, and every property G: if K shares G with its tokens, then K kinstantiates G. Additionally, the reason why (7) seems to us to be true is that we ordinarily use it to say not that K instantiates G, but that K kinstantiates G.[42]

If (Kinstantiation Doctrine) is true, then (i) ought to be handled with care: according to (Kinstantiation Doctrine), a type K and its tokens *share* properties *only* in the sense that K kinstantiates properties that K-tokens instantiate. Importantly, though, (Kinstantiation Doctrine) allows the realist to analyse (5) satisfactorily. Let us suppose that (5) is ambiguous between two statements:

---

42. Importantly, the advocate of (Kinstantiation Doctrine) does not believe that types do not instantiate any properties. According to her, a type instantiates, for instance, the property of being abstract, the property of being a type, and the property of being universal. The properties that a type kinstantiates are only those that it shares with its tokens.

20.   The seriema instantiates the property of being a large bird, and the property of inhabiting the South American forests.

21.   The seriema kinstantiates the property of being a large bird, and the property of inhabiting the South American forests.

(20), we have seen, is untrue: the seriema-type does not instantiate the properties that are attributed to it in (20). (21), by contrast is true. The seriema-type shares with its tokens the properties that attributed to it in (21). By (Kinstantiation Doctrine), the seriema-type kinstantiates these properties, and, hence, (21) is true. If (5) is ambiguous between (20) and (21), then perhaps the reason why (5) seems to us to be true is that we use it in order to express (21) rather than (20).

If my arguments above are sound, then the realist's analysis of (5) is satisfactory: on her view, (5) is true—just as our intuitions say it is. What about (8) and (12), though? Is the realist's analyses of these sentences satisfactory?

I believe that it is. Let us consider (8) first. The type-realist believes that the expression 'the phone [a]' in (8) refers to the [a]-type. Types, by definition, are capable of having instances—or, in other words, occurrences. Moreover, the [a]-type has a plenitude of occurrences, and exactly four of them exist in each and every utterance of the name 'Axayácatl'. In other words, therefore, the [a]-type occurs four times in every utterance of 'Axayácatl'. Thus, the realist's analysis of (8) is undoubtedly satisfactory: on her view, (8) is true as our intuitions suggest—and it is literally so.

What about (12)? On the realist's view, we have seen, 'Old Glory' in (12) refers to a type. Types are, by definition, abstract: they exist outside of spacetime. Since types exist outside of spacetime, a type cannot instantiate distinct properties at distinct points in time. If this is true, then the type theorist cannot understand (12) literally: it cannot be the case that Old Glory instantiated, in 1846, the property of being twenty-eight-starred, and, nowadays, the property of being fifty-starred. In what follows, though, I will show that for decades the realist has possessed the tools required to analyse (12) satisfactorily.

In his (1980, p. 56), Wolterstorff noted that some types are associated with a norms of various kinds. He called them 'norm-types'. A norm-type is a type that may have norm-conforming tokens, and non-norm-conforming tokens. For example: musical works, on Wolterstorff's view, are types and, more specifically, norm-types tokened by perfor-

mances. The norms that musical works are associated with, moreover, are standards of correctness for their performances. Accordingly, a norm-conforming token of a musical work W is a correct performance of W; a non-norm-confirming token, an incorrect performance. Wolterstorff maintained, additionally, that norm-types possess normative properties. A normative property is a property designated by a propositional function of the following form: 'for every y, y cannot be a norm-conforming token of x if y is not G'. Sibelius's *Finlandia*, for example, possesses the normative property designated by the following propositional function: 'for every y, y cannot be a correct token of *Finlandia* if y does not begin with a C-sharp minor chord'. Importantly, moreover, Wolterstorff claimed that attributions of normative properties to norm-types are virtually always imprecise. Witness:

22.   *Finlandia* starts with a C-sharp minor chord.

(22) is intuitively true. If it is true indeed, then it cannot be understood literally: *Finlandia*, we have seen, is a type; and, since types are not located in spacetime, they have no spatiotemporal parts. Consequently, we cannot truly say of a type that it starts with a C-sharp minor chord. (22), therefore, ought to be understood as expressing the following:

23.   *Finlandia* is such that, for all y, y cannot be a correct token of *Finlandia* if y does not start with a C-sharp minor chord.

The reason why we assert (23) instead of (22) is that (22) is simpler and more economical, and, thus, more easily assertible than its more precise counterpart.

The realist's view holds that Old Glory is a norm-type like *Finlandia*. The norms that musical works are associated with, we have seen, are standards of correctness for their performances. The norm that Old Glory is associated with, by contrast, is a condition for Old-Glory-tokens to be law-conforming. More specifically: the norm in question states that, for all y, y cannot be a law-conforming Old-Glory-token at t if the number of stars in y is not identical to the number of States that pertain to the Union at t. The norm that Old Glory is associated with, then, establishes what a norm-conforming Old-Glory-token is in terms of the law of the United States. Since this law changes over time, there are no law-conforming Old-Glory-tokens *simpliciter*, but only Old-Glory-tokens that are law-conforming relative to a certain point in time.

Now, on the type-theorist's view, (12) is nothing more than a imprecise way of attributing to Old Glory the normative property that characterises it. The realist, that is to say, believes that (12) is to be understood as expressing the following:

24. Old Glory is such that, for all y, y cannot be a law-conforming token of Old Glory at time t if the number of stars in y is not identical to the number of States that pertain to the Union at t. Moreover, the number of States that pertained to the Union in 1846 were twenty-eight; today, this number is fifty.

The reason why we assert (12) instead of (24) is, of course, that (12), like (22), is simpler, more economical, and more easily assertible than its more precise counterpart.

If my arguments above are sound, then the realist is able to meet the nominalist's challenge: she is able to account satisfactorily for (5), (8) and (12). Nevertheless, the nominalist could still protest against realism. I have shown that the realist uses three distinct strategies to analyse various t-sentences: some t-sentences (e.g., (8)) she analyses literally; some others (e.g., (5)), she analyses in accordance with (Kinstantiation Doctrine); and others still (e.g., (12)), she analyses in light of the fact that norm-types are often attributed normative properties non-literally. These three strategies allowed the realist to analyse (5), (8) and (12) satisfactorily—and, more specifically: they allowed her to construe these sentences in accordance with our intuitions that they are true.

Nevertheless, the nominalist could contend, it has not been demonstrated that the three strategies in question suffice to construe all t-sentences as true. The possibility exists, therefore, that there is a t-sentence s such that s cannot be construed as true by way of any one of the aforementioned strategies. s, then, is untrue under its literal reading and its kinstantiation reading, and it cannot be adequately understood as a sentence where a norm-type is non-literally attributed a normative property. If this is the case, the nominalist could claim, then the thesis is a mere article of faith that all t-sentences can be construed as true within the realist's framework—or, in other words, under the assumption that they all concern types. Thus, the nominalist concludes, realism is just as untenable as nominalism; realism, like nominalism, is nothing more than an article of faith.

The aforestated objection is unsound, however. T-sentences, by definition, are intu-

itively true sentences that appear to involve reference to types. Thus, by definition, t-sentences involve, under a face-value interpretation, reference to types. The nominalist is committed to an article of faith because she is committed to the thesis that, for every t-sentence s, there is an interpretation of s under which s does not concern types at all. Since the nominalist has not proved that such an interpretation exists for every t-sentence, the nominalist thesis in question is nothing other than an article of faith. There is, however, no analogous article of faith that the realist is committed to. The thesis that there is an interpretation (viz. the face-value interpretation) under which t-sentences concern types is not an article of faith: it is a definitional truth.

At this point the nominalist could protest as follows: the realist has not demonstrated that the three aforementioned strategies suffice to construe all t-sentences as true. Consequently, the realist's belief that these sentences are all true is nothing more than an article of faith. If this is the nominalist's complaint, however, then that which she is now calling an article of faith is nothing other that the intuitive conviction that t-sentences are true. It is important, though, to highlight a certain fact which relates to this conviction. T-sentences seem to involve reference to types. The only reason why we are debating whether types exist is that these sentences are intuitively true. If they were not intuitively true, then there would be no reason whatsoever to suppose that types exist. The conviction that t-sentences are true, then, cannot be adequately characterised as an article of faith; it ought to be characterised, instead, as a conviction that we have all decided to take on as a *datum* in order to determine what there is—in order to determine, more specifically, whether types exist.

If all of the above is right, then there are two things that the realist upholds and are not articles of faith. First: the thesis that, for every t-sentence s, there is an interpretation of s under which s concerns types. And second: the conviction that all t-sentences are intuitively true. It appears, however, that the article of faith that the nominalist initially imputed to the realist corresponds to a slightly, albeit importantly different thesis. The article of faith that the nominalist initially imputed to the realist is the thesis that all t-sentences can be construed as being *both* true and about types. It is an error, though, to suppose that this thesis is an article of faith.

It is true that the realist has not demonstrated that the three aforementioned strategies suffice to construe all t-sentences as true—but she is not committed to the view that they do. She knows, that is to say, that it is possible that a t-sentence s exist which is untrue

under its literal and its kinstantiation readings, and cannot be adequately understood as a sentence where a norm-type is non-literally attributed a normative property. Let K_1 and G_1 be a type and a property respectively, and let s be the following sentence:

25.   The K_1 is G_1.

By assumption, s is untrue under its literal reading, so K_1 does not instantiate G_1. Also by assumption, s is untrue under its kinstantiation reading, so K_1 does not kinstantiate G_1. Finally, s, by assumption, cannot be adequately understood as saying non-literally that a certain norm-type bears a certain normative property. If the realist wants to construe s as true within her framework, she will have to articulate a compelling explanation for why the type K_1 can be nonetheless truthfully said to be G_1. She has not yet done so, though, and, consequently, the nominalist maintains that the realist is committed to an article of faith: she believes that all t-sentences can be construed as being both true and about types—but she does so despite the fact that she has not yet thus construed s.

My answer to the nominalist is this: the realist believes that s can be construed as being both true and about types—but that is not an act of faith at all. Her belief, instead, is wholly rational. It is certain that there exists a time t such that, at t, the only reading that exists of s is its face-value reading—a reading, that is to say, under which s is about types. (If s has been nominalised, then t predates the moment of s's nominalisation; if it has not been nominalised, then t is the present.) By assumption, s is intuitively true. Consequently, at t, s appeared to us to be, at once, both true and about types—and there was no hint of irrationality in our having the impression in question. If s has been nominalised, then we are free to prefer its nominalist reading over the face-value alternative—and we are free to do so in the name of ontological parsimony or something else. (This even the realist admits: she admits, recall, that the nominalist reading of some t-sentences—e.g., (16)—might be considered best.) Even if this is so, however, the fact stands that s *can* be rationally thought to be both true and about types. In simple terms: s is intuitively true, and s, under a face-value reading, refers to types, so it is entirely reasonable to suppose that s is both true and about types. It ultimately follows that it is wholly within the sphere of reason that the realist holds the conviction that 'the K_1' in s refers to a type, and that this type is such that can be *somehow* truthfully said to be G_1. (Perhaps, for instance, the predicate 'G_1' in s is used non-literally to designate not G_1, but a related property. (23) and (24) in

Chapter Five are instances of this.)

A final protest from the nominalist goes as follows: it is true that s, if read at face value, appears to involve reference to a type; but it is also true that, under a face-value reading, what s says of K_1 is that it instantiates G_1. Nevertheless, we concluded earlier that K_1 does not instantiate G_1. (If it did, I noted, s could be adequately interpreted literally, but, by assumption, it cannot.) If the realist is happy to betray our intuitions that what s says of K_1 is that it instantiates G_1, we may just as well betray our intuitions that s involves reference to the type K_1, and opt for a nominalist interpretation of s.

The nominalist is right that s, under a face value reading, says of K_1 that it instantiates G_1—but her protest is futile. For one thing, if she wants us to opt for the nominalist interpretation of s, she must tell us what this interpretation is; she must, in other words, nominalise s—and the burden is on her to show us that s can be effectively nominalised. Additionally, and more importantly, nominalising s *per se* would be largely futile: as I argued in §2, as long as t-sentences are not *all of them* nominalised, we will remain committed to the existence of types.

If my arguments above are sound, then the realist is, *pace* the nominalist, not committed to an article of faith. Additionally, I argued earlier, she is capable of accounting satisfactorily for the same sentences as the nominalist—i.e., (5), (8) and (12). If all of the above is right, therefore, then realism is significantly satisfactory.

4   Conclusion

I devoted this chapter to the twofold task of defining types and defending their existence. The first part thereof I carried out in §1. In §2, I argued that nominalism, the view that types do not exist, is rationally untenable. Finally, in §3, I defended the contention that type realism is a significantly satisfactory theory. If my arguments in §2 and §3 are sound, therefore, then the existence of types is robustly supported. In the following chapter, I will rely on this conclusion in order to advance an original theory about the ontology of *ficta:* a theory that states that *ficta* are best identified with types.

Chapter Five:
The Type Theory

Fictional realism is the view that there are fictional entities. That fictional entities do not exist appears platitudinous, but fictional realism is nonetheless robustly motivated. Witness:

1.   Ireneo is a hypermnesiac.
2.   Ireneo is a Uruguayan man.
3.   Ireneo was created by Jorge Luis Borges in 1942.
4.   Ireneo is one of Borges's most eminent creations.
5.   Some fictional entities are more realistic than others.

(1)–(5) are intuitively true. If they are true indeed, then the name 'Ireneo' refers, and the existential quantifier in (5) ranges over a domain that contains fictional entities. Thus, the fictional realist concludes, our commitment to the truth of (1)–(5) commits us to the existence of *ficta*.

If fictional entities exist, they must belong to a particular ontological category—but which? In Chapter Two and Chapter Three, I argued against two prominent theories of the ontology of *ficta:* possibilism and creationism. If my arguments are sound, then both of these theories are untenable. In this chapter, I will present a novel alternative: the type theory.[43]

In §1, I will advance a set of *desiderata* for a satisfactory account of the ontology of *ficta*. In §2, I will reconstruct a theory of the ontology of *ficta* that is unique in its theoretical virtues; I will present an objection that defeats it definitively, and then I

---

43. Wolterstorff (1980, pp. 134–197), Zalta (1983, pp. 41–50) and Stokke (2020) have defended theories that are somewhat similar to my type theory: they have maintained that *ficta* are *kinds* or *roles*. My theory, however, I will develop in a direction that is wholly different from theirs.

will introduce the type theory as a judicious reaction to this defeat. In §3, I will develop the type theory in detail, and I will show that it fulfils the *desiderata* that I advanced in §1. Finally, in §4 I will dispel a potential objection, and I will succinctly advance one last argument in favour of the type theory.

1 The Ontology of Fictional Entities

My purpose in this section is to formulate a set of sensible *desiderata* for a satisfactory theory of the ontology of fictional entities. These *desiderata* I will later on appeal to in order to gauge the success of my own theory, and compare this theory with its rivals.

Fictive sentences are those that involve purported reference to *ficta*, or existential quantification over *ficta*. Fictive sentences belong to two distinct kinds: the kind of fictional sentences and the kind of critical sentences. Fictional sentences are sentences where *ficta* are attributed their fictional properties; critical sentences are sentences where *ficta* are attributed their critical properties. For every *fictum a*, *a*'s fictional properties are the properties that characterise *a* in the story where *a* figures. Ireneo's fictional properties, for instance, are, amongst others, the following: being a flesh-and-blood man; being Uruguayan; being hypermnesiac; and suffering from insomnia. (1) and (2), then, are fictional sentences: they are sentences where a *fictum* is attributed its fictional properties. *a*'s critical properties, on the other hand, are the properties that are attributed to *a* when it is observed from the perspective of the literary critic—when it is regarded, that is to say, as an authorial creation that may be aesthetically judged. The following are examples of critical properties: being created by an author at a certain point in time; being well constructed; being realistic; being revolutionary; and being typical or exemplar of a certain literary tradition. (3)–(5), then, are critical sentences.

Fictional sentences are intuitively true: witness, for instance, (1) and (2). Critical sentences, too, are intuitively true: (3)–(5) illustrate that this is so.[44] If critical and fictional sentences are true indeed, then *ficta* instantiate both their fictional properties and their critical properties. Nevertheless, extant theorists of the ontology of fictional

---

44. There are sentences that involve purported reference to *ficta*, or existential quantification over *ficta*, but are not intuitively true at all: e.g., 'Ireneo is an old Argentinian man' and 'Ireneo was created by Adolfo Bioy Casares in 1989'. In what follows, though, I will not discuss these sentences, for they are of no interest for those who study the ontology of *ficta*: as I explained earlier, the thesis that fictional entities exist follows from the supposition that certain fictive sentences (e.g., (1)–(5)) are true. Thus, it is the true fictive sentences, and not the untrue ones, that the ontologist of *ficta* is concerned with.

entities have failed to construe these entities as instantiating all of the properties in question. Possibilists, for instance, identify Ireneo with a concretist non-actual individual who instantiates all of Ireneo's fictional properties: he is a flesh-and-blood Uruguayan man, he is a hypermnesiac, and he suffers from insomnia. Nevertheless, the concretist non-actual individual that possibilists identify with Ireneo does not instantiate any of Ireneo's critical properties (Lewis, 1978, p. 38): the individual in question is a flesh-and-blood man, not an authorial creation, and, consequently, he instantiates none of the aesthetic qualities that an authorial creation may bear. Conversely, creationists construe fictional entities as instantiating their critical properties, but not their fictional properties (see, e.g., Von Solodkoff and Woodward, 2017, pp. 407–408). On the creationist's view, Ireneo is an abstract object produced by Borges's authorial endeavours; he is, therefore, an authorial creation that instantiates certain aesthetic qualities. If Ireneo is an abstract artefact, though, then he does not instantiate any one of his fictional properties: evidently, an abstract artefact cannot instantiate the property of being a Uruguayan man; he cannot be hypermnesiac, and cannot suffer from insomnia.

Both creationists and possibilists, then, have failed in construing fictional entities as instantiating both their critical properties and their fictional properties. I think, however, that they cannot be blamed for this failure: it seems to me that there are compelling reasons to believe that it is impossible to construe fictional entities as instantiating all of the properties in question. If Ireneo instantiates both his critical properties and his fictional properties, then he instantiates sheerly conflicting qualities: he is both a flesh-and-blood man and a product of Borges's authorial endeavours; he is both an authorial creation and a man who suffers from insomnia. Those who wish to construe *ficta* as instantiating both their fictional and their critical properties must, therefore, find an ontological category whose members are capable of instantiating such conflicting qualities as those that I just attributed to Ireneo, and then they must identify *ficta* with its members. It appears, though, that no such ontological category exists: the authorial endeavours of writers simply do not produce flesh-and-blood men; the products of the authorial endeavours of writers simply do not suffer from insomnia. If this is right, then there is no ontological category such that, if *ficta* are identified with its members, then *ficta* can be construed as instantiating both their fictional and their critical properties. Theorists of the ontology of *ficta* are thus doomed to fail, like possibilists and creationists, in construing *ficta* as instantiating all of the properties in question.

For every theory T, if *ficta* in T do not instantiate both their fictional and their critical

properties, then there is a compelling argument for the conclusion that T is defective. Suppose that *ficta* in T instantiate their fictional properties, but not their critical properties. Then, within the framework of T, critical sentences are untrue: critical sentences, recall, are those where *ficta* are attributed their critical properties; if *ficta* do not instantiate their critical properties, then critical sentences are untrue. Now suppose that *ficta* in T instantiate their critical properties, but not their fictional properties. Within the framework of T, then, fictional sentences are untrue: I said earlier that fictional sentences are those where *ficta* are attributed their fictional properties; if *ficta* do not instantiate their fictional properties, then fictional sentences are untrue. In sum: for every theory T, if *ficta* in T instantiate their fictional properties or their critical properties, but not both, then T entails that either critical sentences or fictional sentences are untrue. Earlier, however, I argued that it is not possible to formulate a theory of the ontology of fictional entities that construes these entities as instantiating both their critical and their fictional properties. If my argument was sound, then all the theories of the ontology of *ficta* are bound to construe either fictional or critical sentences as untrue.

Importantly, any theory that construes fictional or critical sentences as untrue is a theory that compels us, on pain of irrationality, to revise our discursive practices: we often advance critical and fictional sentences, but advancing untrue sentences is irrational. Revising our discursive practices in the required manner is undesirable, though, for both critical and fictional sentences constitute fundamental means for us to express our thoughts concerning literature: they are crucial for us to describe and discuss the contents of works of fiction (cf. (1) and (2)), and to do literary criticism (cf. (3)–(5)). Any theory that construes fictional or critical sentences as untrue, therefore, delivers an undesirable species of revisionism.

Since the revision of our discursive practices is undesirable for the aforestated reasons, a satisfactory theory T of the ontology of *ficta* is one that fulfils the following *desideratum:*

> **Anti-Revisionism**
> The truth of T is consistent with the thesis that we speak truly when we assert both fictional sentences and critical sentences.

Earlier I presented an argument to the effect that all theories of the ontology of fictional entities are bound to construe either critical or fictional sentences as untrue. If this argument is sound, then there does not exist a theory that fulfils (Anti-Revisionism).

Fortunately, there is a way for theorists of the ontology of *ficta* to contend that the argument in question is unsound: they need only claim that we have been misunderstanding fictive sentences all along.

Suppose that there is a theory T such that, according to T, *ficta* instantiate their fictional properties, but not their critical properties. Within the framework of T, fictional sentences are true, but it appears that critical sentences are straightforwardly *un*true. The Tist, though, claims that they are not. According to the Tist, critical sentences are not straightforwardly untrue because they are ambiguous. Consider (3), for instance. The Tist maintains that (3) has two precisifications:

    6.   Ireneo instantiates the property of having been created by Borges in 1942.

    7.   Ireneo is related by relation R to the property of having been created by Borges in 1942.

(6) is untrue in T: by assumption, *ficta* in T do not instantiate their critical properties; thus Ireneo, in T, does not instantiate the property of having been created by Borges in 1942. However, the Tist claims, (7) is true: R is a relation that is different from instantiation, but does in fact relate Ireneo and the property in question. Moreover, the Tist contends, we speak truly when we assert (3) because we use it to express (7) rather than (6).

In more general terms: within the framework of T, we do speak truly when we assert critical sentences. We also speak truly when we assert fictional sentences, for fictional sentences are, by assumption, true in T. Consequently, T does fulfil (Anti-Revisionism).

As it happens, the creationist, with the aim of abiding by (Anti-Revisionism), makes use of a strategy that is exactly analogous to that of the Tist. The creationist, we have seen, identifies *ficta* with abstract artefacts. On her view, *ficta* instantiate their critical properties, but not their fictional properties: Ireneo, for instance, does not instantiate the property of being a Uruguayan man. And yet, the creationist notes, it is clear that Ireneo is related to said property by a special relation—a relation that he does not bear to, say, the property of being an old Argentinian man. Hence, even if *ficta* do not instantiate their fictional properties, there exists a relation between *ficta* and these

properties.[45] The creationist decided to call this relation 'holding'.[46] Additionally, the creationist contends that fictional sentences are ambiguous: they can be read as stating that *ficta* instantiate their fictional properties, or as stating that *ficta* hold these properties. On their view, (2), for example, is ambiguous between (8) and (9):

8. Ireneo instantiates the property of being a Uruguayan man.
9. Ireneo holds the property of being a Uruguayan man.

(8), the creationist says, is false, but (9) is true, and we speak truly when we assert (2) because we use it to express (9) rather than (8) (see, e.g., Van Inwagen, 1983, pp. 75–76). In more general terms: within the creationist framework, we do speak truly when we assert fictional sentences. Since *ficta*, within the creationist framework, bear their critical properties, then critical sentences are true within said framework. Thus, the creationist concludes, creationism fulfils (Anti-Revisionism).

In sum: there is a strategy that allows any theorist of the ontology of *ficta* to comply with (Anti-Revisionism): she need only contend that the sentences that she struggles to handle (be them fictional or critical sentences) have not only a reading that is untrue within the framework of her theory, but a true reading also. The postulation of an ambiguity, however, is a move that must be executed with caution. As Kripke (1977, p. 269) notes, postulating an ambiguity often suffices to vindicate a theory in the face of a piece of putative counterevidence—but this is so even when the theory in question is false. For instance: suppose that the tripartite analysis of knowledge is false—as many epistemologists nowadays believe it to be. Suppose, moreover, that an individual *a* is convinced that the tripartite analysis is true. At some point, *a* is presented with Gettier's counterexamples: hypothetical scenarios where, intuitively, a subject possesses justified true belief, but not knowledge. *a* dismisses the counterexamples: we are right, she maintains, to judge that the subjects that star Gettier's scenarios possess justified true belief; but only because, when we so judge them, we use the term 'justified' in a special sense—a sense that is distinct from the sense that 'justified' has when it occurs in the statement that knowledge is justified true belief. Consequently, *a* maintains, there are no true counterexamples to the tripartite analysis of knowledge; the tripartite analysis, that is to say, remains undefeated.

---

45. This argument is originally from Van Inwagen (1983, p. 75).

46. The term 'holding' was originally coined by Van Inwagen (1983, p. 75).

Given the ease with which even a false theory might be vindicated by the postulation of an ambiguity, Kripke (1977, p. 269) advances the following cautionary principle: an ambiguity must be postulated only if we are inescapably forced to do so. Kripke's principle is grounded on the fundamental principles of theoretical inquiry: it is grounded on the principle of parsimony, for it entails that senses ought not to be multiplied beyond necessity; and it is grounded on the principle of sufficient reason, for it entails that senses may be posited only when we have a compelling reason to do so. Consequently, I believe that Kripke's principle must be abided by in the course of our inquiry concerning the ontology of *ficta*. I believe, in other words, that a satisfactory theory T of the ontology of *ficta* fulfils not only (Anti-Revisionism), but the following *desideratum* as well:

> **Semantic Parsimony**
>
> T does not involve the postulation of an ambiguity that we are not strictly compelled to postulate.

Sadly, (Semantic Parsimony) and (Anti-Revisionism) appear to be at odds. Earlier I argued that theorists of the ontology of *ficta* are bound to construe *ficta* as not instantiating both their critical and their fictional properties. This, I claimed, will lead these theorists to construe either fictional or critical sentences as untrue; and this, in turn, will lead them to adopt the revisionist stance that (Anti-Revisionism) proscribes—unless, I said, they contend that the relevant sentences are ambiguous. Nevertheless, (Semantic Parsimony) imposes very strict limits to our abilities to postulate ambiguities. How, then, could (Semantic Parsimony) and (Anti-Revisionism) ever be both fulfilled?

In what follows, I will advance a theory that fulfils them both: the type theory. The type theory involves appeals to not one, but several ambiguities that serve the purpose of providing both fictional and critical sentences with true readings. Consequently, they serve the purpose of turning the type theory into a theory that fulfils (Anti-Revisionism). Nevertheless, all of the ambiguities that the type theory involves are ambiguities that are strongly supported, and supported independently of the interests of the type theorist: they were posited years ago, and will continue to be upheld even if the type theory is rejected. The type theorist abides by (Semantic Parsimony), then, for two independent reasons. First, she does not really postulate any ambiguity: she merely appeals to ambiguities that are already there. And, second: as I will show below, there are compelling reasons to believe that the ambiguities in question ought to be posited in order to make sense of a very large portion of our everyday assertions,

and an incalculable amount of scientific statements. In sum: as Kripke (1977, p. 267) suggests, the best ambiguities for one to endorse for a given purpose are those which are needed anyway, and those are precisely the ambiguities that the type theorist endorses.

2   The Problem of the Multiple Candidates

In this section, I will reconstruct a theory of the ontology of *ficta* that is unique in its theoretical virtues, and I will present an objection that defeats it definitively. Then I will introduce the type theory as a judicious reaction to this defeat.

In §1, I said that, for every *fictum a*, the properties that characterise *a* in the story where *a* figures are *a*'s fictional properties. I also noted that our intuitions tell us that we speak truly when we attribute *a*'s fictional properties to *a*—recall, for example, (1) and (2). Several of the extant theories of the ontology of *ficta* fail to construe these entities as instantiating their fictional properties. For example: creationism, which I mentioned in the previous section, is the view that fictional entities are abstract artefacts. On the creationist's view, then, Ireneo is abstract, and, consequently, he does not instantiate any one of his fictional properties: as we saw earlier, a *fictum* cannot instantiate the property of being a flesh-and-blood man; it cannot be a hypermnesiac, and it cannot suffer from insomnia.[47]

Possibilists, by contrast, do succeed in construing fictional entities as instantiating fictional properties. Possibilism, we saw in §1, is the view that *ficta* are concretist non-actual individuals. The concretist non-actual individuals that possibilists countenance, however, are infinitely many and vastly diverse. If *a* is a *fictum*, which one of those concretist non-actual individuals is the one that possibilists identify with *a*? The answer to this question is contained in the following principle, which I mentioned in passing in §1:

Possibilist Principle
If *a* is a *fictum*, and $G_1...G_n$ are *a*'s fictional properties, then *a* is identical to the

---

47. Meinongians, too, fail to construe *ficta* as instantiating their fictional properties. Meinongians (see, e.g., Parsons, 1980; Zalta, 1983; Jacquette, 1996) identify fictional entities with Meinongian objects—with objects, that is to say, that do not exist. According to the metaphysics that underlies the semantics of first-order logic, however, instantiation without existence is definitionally impossible: a property G is instantiated just in case there exists an x such that x instantiates G. Consequently, Meinongians face difficulties in attributing to *ficta* their fictional properties. See Kroon (2022) for a summary of the Meinongian's various attempts to solve these difficulties.

concretist non-actual individual who possesses $G_1...G_n$.

(Possibilist Principle) entails trivially that, for every *fictum a*, *a* bears *a*'s fictional properties. Within the possibilist framework, then, a statement where *a* is attributed its fictional properties is a true statement—and that is exactly what our intuitions tell us: (1) and (2), for example, are statements where a *fictum* is attributed its fictional properties and, intuitively, both (1) and (2) are true.

Possibilism thus possesses a theoretical virtue that its extant rivals lack: it fosters a perfectly simple account of the truth of fictional sentences, whilst the advocates of other theories ought to resort to rather non-parsimonious doctrines in order to do so—recall, for instance, the ambiguity thesis proposed by creationists and described in §1.

Unfortunately, possibilism is subject to an objection that appears unsurmountable. Kripke (1980, pp. 156–158; 2013, pp. 41–53) observed that, in logical space, there exist not one, but multiple non-actual individuals that instantiate each and every one of Ireneo's fictional properties. Each one of these individuals, then, resembles Borges's protagonist as much as the others. Consequently, Kripke maintains, for each one of the individuals in question, we are radically precluded from establishing that him, and not any one of the others, is Ireneo. Thus Kripke contends that Ireneo cannot be identified with a non-actual individual. If we generalise this contention in the obvious manner to the totality of fictional entities, we will obtain the conclusion that possibilism is untrue: *ficta* are not non-actual individuals at all.[48]

Kripke's objection consists, in simple terms, in pointing out that possibilists face a problem of multiple candidates: there are multiple Ireneo-candidates, but none of them can be elected as the one that Ireneo is identical to. If we wish to take advantage of the extraordinary theoretical virtues possessed by possibilism, we ought to solve the problem in question; and, in order to articulate such a solution, we may want to learn how analogous problems have been solved in the past.

In the foundations of mathematics, set-theoretic foundationalism is the view that mathematical objects are best identified with sets. Benacerraf argued against this view in his (1965). There are multiple candidate set-theoretic reductions of the natural numbers—multiple systems of set-theoretic objects, that is to say, that the natural numbers

---

48. See §1 in Chapter Three for a detailed reconstruction of this objection.

could be identified with. They could be identified with the finite von Neumann ordinals, with the finite Zermelo ordinals or, indeed, with any one of the members of an infinitely large collection of set-theoretic systems that are isomorphic to the two that I just mentioned. If the natural numbers are sets, as set-theoretic foundationalists maintain, they must be identical to exactly one of these systems—but which? According to Benacerraf, for every one of these systems, there is no reason whatsoever to suppose that the natural numbers are identical to it and not any one of the others. It follows that, for each one of the systems that we are considering, we cannot establish that it, and not any one of the others, is identical to the natural numbers (1965, p. 62). Benacerraf, in sum, showed that set-theoretic foundationalists face a problem of multiple candidates: there exist multiple candidate set-theoretic reductions of the natural numbers, but none of them may be elected as the one that the natural numbers are identical to.

In order to solve this problem, Benacerraf (1965, pp. 69–73) proposed to identify the natural numbers not with any of the set-theoretic systems that I mentioned above, but with their *structure:* with a universal entity, that is to say, *exemplified* by all of the systems in question.[49] This structure is unique: no other entity of its kind is exemplified by the systems that we have been considering.[50] Since it belongs to a higher ontological order, moreover, it is positively distinguished from the systems that exemplify it.[51] Consequently, we have a principled reason to identify the natural numbers with it instead of its exemplifications. And, if we do so, we will have solved Benacerraf's problem of multiple candidates: there will be one particular entity that the natural numbers are identical to.

Benacerraf's proposal was later carried to several areas of mathematics, and developed in various directions (see, e.g., Hellman, 1996; Resnik, 1997; Shapiro, 1997). None of that, however, is relevant now: I have presented Benacerraf's proposal only in order

---

49. The term 'exemplified' here has the meaning that I assigned it in §1 of Chapter Four and, more precisely, in (Exemplification). The entity in question is not a type, but that is irrelevant for our present purposes. See, e.g., Hellman (1996); Resnik (1997); Shapiro (1997) for various developments of Benacerraf's proposal.

50. Or so Benacerraf thought. See McGee (1997, p. 151); Hellman (2001, pp. 193–196; 2005, p. 546); Button & Walsh (2016, pp. 287–288); Assadian (2018) for arguments to the effect that Benacerraf's structure is in fact not unique.

51. Belonging to a higher-order ontological category has been thought to distinguish certain universal entities positively from their exemplifications for reasons that I do not have the space to discuss here, but are very clearly presented by Leitgeb (2022, pp. 323–328).

to suggest that it may harbour an insight for those who wish to solve the possibilist's problem.

Benacerraf's solution to his problem of multiple candidates consists, in sum, in identifying the natural numbers with an abstract entity exemplified by the candidates he could not choose from. And the possibilist may do the same: she may, that is to say, identify Ireneo with an abstract entity exemplified by all of her Ireneo-candidates.[52] This is precisely what the type theorist does: she identifies Ireneo with the type tokened by all of the Ireneo-candidates that the possibilist countenances. This type is unique: there is no other type exemplified by the non-actual men in question. And, since it belongs to an ontological order higher than that of its tokens, there is a reason to distinguish it positively from them. Thus we have a principled reason to identify Ireneo with this type instead of any one of its tokens; and, if we do so, we will have solved Kripke's problem of multiple candidates: there will be one particular entity that Ireneo is identical to.

3   The Type Theory

In this section, I will develop the type theory in detail; I will show that it fulfils the *desiderata* that I advanced in §1, and I will demonstrate that it possesses a number of desirable theoretical virtues.

According to the type theory of the ontology of fictional entities, fictional entities are types. In particular:

> **Axiom 1**
>
> For every *fictum a*, *a* is the type tokened by all of the non-actual individuals that instantiate *a*'s fictional properties.[53]

---

52. An alternative solution to the possibilist's problem has been suggested independently by Caplan (2016) and Haraldsen (2017; 2020). Their suggestion, simply put, is that the possibilist appeals to the theory of arbitrary reference (Breckenridge & Magidor, 2012) in order to latch the name 'Ireneo' onto one of her Ireneo-candidates. In Chapter Three, however, I argued that this suggestion fails.

53. At this point, the question may arise of what determines which properties are *a*'s fictional properties. I believe that this question deserves its own attention, and I lack the space to discuss it properly here. Nevertheless, it may be answered tentatively as follows: which properties are *a*'s fictional properties is determined by our interpretation of the story where *a* figures. Consider, for instance, *Funes, the Memorious*. If we imaginatively engage with *Funes, the Memorious*, and we interpret it correctly, it will become clear to us that we are supposed to imagine an individual as bearing certain properties: we are supposed to imagine an individual called 'Ireneo'; we are supposed to imagine that he is a Uruguayan man, that he is hypermnesiac, etc. We may imagine him also as having a scar on his chin—but we know that we need not do so. I think, therefore, that the method that allows us to determine which properties

Hence, on the type theorist's view, Ireneo, for example, is the type instantiated by all of the Ireneo-candidates that I talked about in the previous section: all of the non-actual men, that is to say, who instantiate Ireneo's fictional properties.

A semantically inflated paraphrase of (Axiom 1) says that fictive names refer to types. If fictive names refer to types, then the things we read about when we read a work of fiction are types. This result might appear counterintuitive: intuitively, when we read of Ireneo we read of a flesh-and-blood man, not an abstract entity. The counterintuitiveness of the result in question, however, vanishes when we recall Wollheim's characterisation of types: we think, he says, of a type as if it were one of its archetypical tokens. If he is right, then the thing that we read about when we read Borges's story is an abstract entity, but we imagine it as if it were an archetypical Ireneo-candidate—as if it were, that is to say, a flesh-and-blood Uruguayan man who is a hypermnesiac, suffers from insomnia, etc.

To my mind, this type-theoretic proposal serves to solve one of the most perplexing puzzles that exist concerning fiction. It appears unquestionable that the mental images that we produce when we read Borges's story are images of a man—but which man? Not an actual one, of course: the man that we imagine we might imagine as resembling a person that is familiar to us (e.g., an nephew), but it is clearly not that person that we imagine as having Borges's adventures. It seems, moreover, that the man that we imagine is not non-actual either. As Kripke (1980, pp. 156–158; 2013, pp. 41–53) noted, there are multiple non-actual men who resemble Ireneo precisely, but we are radically incapable of selecting any one of these men amongst the others: any one description that we could ever produce to try to select one of them will be satisfied by many. Hence, selecting one of these men is effectively impossible. If this is the case, then we could have never singled out one of the men in question in order to imagine *him* as we read Borges's story. If the man that we imagine is not actual nor non-actual, then what is it? The type theorist's response now appears exactly adequate: we imagine not any one of the men who resemble Ireneo, but the type thereof: we imagine, that is to say, an *abstractum* that appears in our minds as an archetypical Ireneo-like man.

---

are a's fictional properties may be hermeneutical: if we interpret a's story correctly, it will become clear to us that we are supposed to imagine a as bearing certain properties—and those properties, I believe, are a's fictional properties. (A similar proposal was endorsed by Von Solodkoff and Woodward, 2017, p. 412.)

Now, in order to further motivate the type theory, I will show that it fulfils the *desiderata* that I proposed in §1.

In §2 of Chapter Four, I contended, following Wollheim (1980, pp. 50–51), that, for every type K, K shares properties with its tokens. The types that the type theorist identifies with *ficta* are exactly like K; and, more specifically:

Axiom 2

For every *fictum a*, a shares a's fictional properties with its tokens.

The following is a trivial entailment of (Axiom 2):

Theorem 1

For every *fictum a*, a possesses a's fictional properties.

The first of my *desiderata* is (Anti-Revisionism). It states that a satisfactory account of the ontology of *ficta* vindicates our conviction that we speak truthfully when we assert both fictional sentences and critical sentences. Let us first investigate whether the type theory vindicates our convictions concerning fictional sentences.

Fictional sentences, we have seen, are those where *ficta* are attributed their fictional properties. (Theorem 1) states that *ficta* possess their fictional properties. Thus it follows that fictional sentences are true within the framework of the type theory. Consider, for example, (1) and (2). By (Theorem 1), Ireneo possesses Ireneo's fictional properties: he is, then, amongst other things, a hypermnesiac Uruguayan man. Thus (1) and (2) are both true.

In fact, however, this conclusion requires a small qualification. In §3 of Chapter Four, I noted that types are often unfit to instantiate some of the properties that they share with their tokens. Then I argued, accordingly, that the properties that a type K shares with its tokens are possessed by K in a peculiar sense: K and the properties in question are related not by instantiation, but kinstantiation. I also argued in §3 of Chapter Four that there is an ambiguity in the sentences where types are attributed the properties that they share with their tokens. More specifically: I proposed, by way of (Kinstantiation Doctrine), that (10) is ambiguous between (11) and (12):

10.   The K is G.
11.   The K instantiates G.

12.   The K kinstantiates G.

(11), I contended, is false, but (12) is true and, if (10) seems to us to be true, it is because we use it to express (12) rather than (11).

The types that the type theorist identifies with *ficta* are not exceptions to any of the above. First, these types appear unfit to instantiate the properties that they share with their tokens. Ireneo, for instance, is, according to the type theorist, the type tokened by all of the non-actual individuals that instantiate Ireneo's fictional properties. By (Axiom 2), Ireneo shares with its tokens Ireneo's fictional properties. Some of Ireneo's fictional properties are the following: being a Uruguayan man, being a hypermnesiac, and suffering from insomnia. Clearly, none of these properties may be instantiated by an abstract object. If Ireneo is a type, as the type theorist maintains, then he ought to be constructed as kinstantiating, rather than instantiating, the properties in question.

Now, fictional sentences are sentences where *ficta* are attributed their fictional properties. On the type theory, every *fictum a* is a type, and it shares *a*'s fictional properties with its tokens. On the type theory, then, fictional sentences are sentences where types are attributed the properties that they share with their tokens. If (Kinstantiation Doctrine) is true, then sentences like these are ambiguous between an instantiation reading and a kinstantiation reading. Consider (2), for example. If my arguments are sound, then (2) is ambiguous between the following two statements:

13.   Ireneo instantiates the property of being a Uruguayan man.
14.   Ireneo kinstantiates the property of being a Uruguayan man.

(13), we have seen, is untrue, but (14) is true; and it might well be that the reason why (2) seems to us to be true is that we use it to express (14) rather than (13).

If the above is true, then (Theorem 1) ought to be qualified: for every *fictum a*, *a* possesses *a*'s fictional properties only in the sense that it kinstantiates them. This suffices, however, to deliver the desired results: if fictional sentences are, as I proposed, ambiguous between an instantiation precisification and a kinstantiation precisification, and we ordinarily intend the latter, then the fact that *ficta* kinstantiate their fictional properties suffices to vindicate our conviction that we speak truthfully when we utter fictional sentences.

It might appear, however, that, by postulating the instantiation-kinstantiation ambi-

guity, I have violated our second *desideratum.* (Semantic Parsimony), recall, states that a satisfactory theory of the ontology of *ficta* does not involve the postulation of an ambiguity that we are not strictly compelled to postulate. An objector to the type theory could contend that we are not strictly compelled to postulate the instantiation-kinstantiation ambiguity. Instead, the objector could claim, we could identify *ficta* with abstract artefacts, as creationists do, and contend that fictional sentences are ambiguous not between an instantiation reading and a kinstantiation reading, but between an instantiation precisification and a holding precisification. The latter ambiguity suffices, just as much as the former, to vindicate our contentions that we speak truthfully when we utter fictional sentences.[54] Consequently, we are not compelled to postulate the instantiation-kinstantiation ambiguity, and we would violate (Semantic Parsimony) if we did so.

In Chapter Two, I argued that the instantiation-holding ambiguity is deeply flawed and, in fact, does not suffice to vindicate our intuitions that we speak truthfully when we utter fictional sentences. Nevertheless, even if my arguments in Chapter Two were unsound, there is a powerful reason to prefer the instantiation-kinstantiation ambiguity over its alternative. Simply put: the instantiation-kinstantiation ambiguity ought to be posited anyway. In §3 of Chapter Four, I showed that this ambiguity must be posited in order to account for the truth of a significant portion of the wealth of the t-sentences that occur in ordinary speech and various scientific discourses. Wetzel (2009, pp. 11–12) offers an alternative account. In Chapter Four, however, I argued that it is not satisfactory. If my arguments are sound, then the instantiation-kinstantiation ambiguity ought to be postulated. And, *pace* my objector, the type theorist does not postulate any ambiguity that she is not strictly compelled to postulate.[55]

Let us now investigate whether the type theory vindicates our intuitions that we speak truthfully when we utter critical sentences. Critical sentences, recall, are sentences where *ficta* are attributed critical properties. The critical properties of a *fictum a* are

---

54. See §1 for a reminder of how the creationist's ambiguity suffices to vindicate our intuitions that we speak truly when we utter fictional sentences.

55. As a matter of fact, I said in §3 in Chapter Four that my arguments against Wetzel's account are not crucial for the success of the type theory. Now it is possible to see why this is so. If Wetzel is right, then types do really instantiate the properties that they share with their tokens—and they do so despite being abstract. Hence, if Wetzel is right, *ficta* do really instantiate their fictional properties. And now the type theorist has no need whatsoever to endorse the instantiation-kinstantiation ambiguity. She has, that is to say, one fewer ambiguity to validate on the face of (Semantic Parsimony). Thus Wetzel's being right would be not a problem, but a benefit for the type theorist.

the properties that are attributed to a when it is observed from the perspective of the literary critic—when it is regarded, that is to say, as an authorial creation that may be aesthetically judged.

In order to vindicate our convictions concerning critical sentences, the type theorist sorts these sentences into two distinct classes. The first comprises sentences that describe *ficta* as authorial creations; the second, sentences where *ficta* are attributed aesthetic properties.

Let us consider the first class first. (3), recall, belongs to this kind: it describes Ireneo as an authorial creation. (3), moreover, is intuitively true—but it cannot be true literally if Ireneo is a type. Under a literal interpretation, (3) says that Ireneo was brought into existence by Borges in 1942. Nevertheless, if Ireneo is a type, then he is abstract and, thus, atemporal; he did not come into existence at any point in time, and, consequently, he was not brought into existence by Borges in 1942. Under a literal interpretation, therefore, (3) is false—but the type theorist can nonetheless vindicate our intuitions that we speak truthfully when we utter (3). Let us see.

Numerous authors (e.g., Kivy, 1983; 1987; 2000; 2002; 2007, pp. 112–142; Haraldsen, 2017) have argued that some of the acts that we intuitively identify as acts of creation are acts of creative selection. For instance: those who adhere to a Platonist doctrine in the philosophy of music identify musical works with universals instantiated by spatiotemporally located sequences of sounds—universals, that is to say, instantiated by performances. On their view, then, musical works are atemporal entities that were not brought into existence by composers at any point in time. The act that we conceive of as the creation of a musical work, then, ought to consist in something else. According to Platonists, it consists in creatively selecting a particular sound-sequence-type amongst the plenitude of sound-sequence-types that exist in Platonic heaven.

The type theorist agrees with the Platonists that some of the acts that we intuitively identify as acts of artistic creation are acts of creative selection. On her view, fictional entities, like musical works, are creatively selected by authors amongst the vast multitude of entities that inhabit Platonic heaven. Thus the type theorist believes that Ireneo, for example, was creatively selected by Borges in 1942. Additionally, the type theorist claims that (3) possesses both of the following readings:

15.   Ireneo was brought into existence by Borges in 1942.

16.   Ireneo was creatively selected by Borges in 1942.

(15), we have seen, is false, but (16) is true, and the reason why (3) seems to us to be true is that we use it to express (16) rather than (15). Thus, the type theorist concludes, our intuitions that we speak truthfully when we utter (3) have been vindicated.

The aforestated type-theoretical thesis concerning (3) generalises to all of those critical sentences that belong to the same class as (3). I said, however, that there are critical sentences that belong to a second class: those where fictional entities are attributed aesthetic properties. Some of these sentences are straightforwardly true within the type theorist's framework. Witness:

17.   John Updike's Fawcett is a more realistic character than Ireneo.

Intuitively, a character is realistic just in case it bears properties that an actual entity could plausibly bear. By (Theorem 1), the type that Ireneo is identical to bears the following properties: having acquired extraordinary cognitive powers after suffering a serious equestrian accident; being radically incapable of oblivion, and having learned Latin in a matter of days. These are not properties that an actual entity could plausibly bear. By contrast, the type that Fawcett is identical to bears, by (Theorem 1), properties like the following: being a middle-aged, American man; being depressed, and being multiply divorced (cf. Updike, 1991). All of these are properties that an actual entity could plausibly bear. In virtue of (Theorem 1), then, the type theorist construes Fawcett and Ireneo as being such that the former is more realistic than the latter. (17) is therefore true within the type-theorist's framework.

Not all of the sentences where *ficta* are attributed aesthetic properties are like (17), however. Suppose that author a wrote a story, F, about a woman called 'b' who has no interest whatsoever in men, marriage or motherhood. Let $G_1...G_n$ be the properties attributed to b in F. Suppose, moreover, that S was written within a literary tradition where female characters are invariably portrayed as women whose lives revolve around motherhood, marriage and men. Critics claim, and justifiably so, that b is a revolutionary character: after the publication of F, a multitude of characters are created of women whose lives do not revolve around men, marriage or motherhood. Suppose now that, decades later, author $c \neq a$ produces a work of fiction that is, by sheer coincidence, lexically identical to F. Let it be called 'F''. Since F' is lexically identical to F, it, too, concerns a woman called 'b' who bears $G_1...G_n$. It follows that the fictional properties

possessed by a's b and those possessed by c's b are exactly the same: $G_1...G_n$.

By (Axiom 1), the type theorist identifies a's b with the type tokened by the non-actual individuals who instantiate $G_1...G_n$. Also by (Axiom 1), she identifies c's b with the type tokened by the non-actual individuals who instantiate $G_1...G_n$. On the type theorist's view, therefore, a's b and c's b are numerically identical. This result is problematic, though. As I said above, critics claim that a's b is a revolutionary character. For obvious reasons, however, they do not judge c's b to be revolutionary in the slightest. If a's b and c's b are numerically identical, how could the following critical statements be both true, as they appear to be?

18. a's b is a revolutionary character.
19. c's b is not a revolutionary character.

An answer to this question may be drawn from Lebens's work. In one of Jorge Luis Borges's short stories, a literary critic celebrates the last of the pieces written by an author called 'Pierre Menard': a novel entitled '*Don Quixote*'. "Cervantes's text and Menard's are verbally identical", says the critic, "but the latter is almost infinitely richer":

> [Cervantes], for instance, wrote (Don Quijote, first part, ninth chapter):
>
> *... truth, whose mother is history, adversary of time, depository of deeds, witness of the past, exemplar and announcer of the present, warning of what is to come.*
>
> Written in the seventeenth century, written by the "ingenious layman" Cervantes, this enumeration of attributes is a mere rhetorical praise of history. Menard, by contrast, writes:
>
> *... truth, whose mother is history, adversary of time, depository of deeds, witness of the past, exemplar and announcer of the present, warning of what is to come.*
>
> History, *mother* of truth; the idea is astounding. Menard, contemporary with William James, defines history not as an inquiry into reality, but as its origin. Historical truth, for him, is not what happened; it is what we judge has happened. The final clauses—exemplar and announcer of the present, warning of what is to come—are shamelessly pragmatistic (Borges, 1944b, pp. 52-53).[56]

There are, Lebens (2015, pp. 3060-3061) notes, weighty reasons to believe that lexically identical works of literature are distinct: as Borges's story illustrates, works of

---

56. The translation from the original in Spanish is mine.

literature that resemble each other exactly may bear different aesthetic qualities. By Leibniz's Law, it follows that verbally identical works cannot be numerically identified. Nevertheless, there are also weighty reasons to believe that lexically identical works of literature are identical also numerically: reading one and them the other is as redundant as reading either one of them twice; by memorising either one of them, one has *ipso facto* memorised the other; by producing a copy of either one of them, one has *ipso facto* produced a copy of the other (Lebens, 2015, pp. 3059–3060).

Lebens thinks that the only way to make sense of these conflicting reasons consists in maintaining that the concept of a work of literature and the expressions that purport to refer to particular works of literature are ambiguous. Suppose that 'W' and 'W'' refer to lexically identical works of literature. Normally, when we speak of W and W', Lebens claims, we use both 'W' and 'W'' to refer to one and the same abstract linguistic structure instantiated by a certain class of strings of ink marks. Let us call it L. Nevertheless, when we attribute inconsistent aesthetic qualities to W and W', we use 'W' and 'W'' to refer not to L, but to two distinct creative processes. One of them is the process by means of which the author of W chose L amongst the plenitude of linguistic structures that exist in platonic heaven, and, by handwriting, by using a typewriter or a word-processing software, produced a concrete, visible token of L in order to turn L into a focus of aesthetic appreciation. The other one is the process by means of which the author of W' chose L amongst other linguistic structures, and produced a concrete, visible token of L in order to turn L into a focus of aesthetic appreciation (2015, p. 2079). What we speak about when we attribute aesthetic properties to a literary work, therefore, is not an abstract entity, but the process by means of which this entity was, so to speak, introduced into the sphere of culture. It is processes like this one that bear contextually determined aesthetic properties such as being revolutionary.

The type theorist contends that the expressions that purport to refer to fictional entities are ambiguous in this way as well. Normally, she claims, when we speak of *ficta*, we speak of types; when we attribute contextually-determined aesthetic properties to *ficta*, however, we speak not of types, but of the creative processes by means of which some types are selected amongst others and turned into a focus of aesthetic appreciation. It is processes like this which we judge as revolutionary. On the type theorist's view, therefore, it is a's creative act of selecting b that is revolutionary, not b itself. When we say that a's b is revolutionary, 'b' refers to said act. When we say that c's b is not revolutionary, 'b' refers to a distinct act: the act whereby c selected b. Hence, the type

110

theorist can say, without absurdity, that both (18) and (19) are true.

In sum: the theory of creation as creative selection, along with Lebens's ambiguity, suffice for the type theorist to account for critical sentences of both of the kinds that exist. If this is the case, then the type theory succeeds in fulfilling (Anti-Revisionism): it is consistent with the thesis that we speak truthfully when we assert both fictional sentences and critical sentences. What about (Semantic Parsimony), though? We might think that, by adhering to the theory of creation as creative selection and endorsing Lebens's ambiguity, the type theorist twofoldly violates (Semantic Parsimony)—she postulates not one, but two ambiguities that we are not compelled to postulate. This thought is mistaken, however: the ambiguities in question both ought to be countenanced. Consider:

20.   Sibelius created the symphonic poem called '*Finlandia*'.

There is a compelling reason to believe that musical works are universal *abstracta*: they admit of multiple performances and, hence, of multiple occurrences. They are, then, not particular, but universal, and all universal entities are abstract. Now, universal *abstracta* are atemporal. If musical works are universal *abstracta*, then they were not brought into existence by composers at any point in time. If we are committed to the truth of (20), then we ought to admit that the verb 'to create' does not always refer to the act of bringing something into existence. If musical works exist atemporally in Platonic heaven, then the relation that exists between their composers and them is one that corresponds to acts of creative selection. If (20) is true, therefore, then the occurrence of 'to create' that it contains refers to one such act. Thus we are compelled to postulate the relevant ambiguity.

Admittedly, the view that musical works are atemporal *abstracta* is not unrivalled. Moruzzi (2018), for example, maintains that musical works are concrete: on her view, the terms that we conceive of as referring to musical works refer to either individual performances or collections thereof. Nevertheless, adhering to this view would not free us from the need to postulate an ambiguity in the verb 'to create'. Each performance of *Finlandia* is constituted by a wealth of spatiotemporally located sounds produced by a multitude of instruments. None of the performances of *Finlandia* is such that Sibelius brought into existence the totality of the sounds that constitute it, so no performance of *Finlandia* was brought into existence by Sibelius. If the term '*Finlandia*' in (20) refers, as Moruzzi says it does, to a performance or a collection thereof, then the occurrence of

the verb 'to create' in (20) does not refer to the act of bringing something into existence. It follows that a further meaning must be countenanced.

Additionally, if we are to abide by (Semantic Parsimony), and refrain from multiplying meanings beyond necessity, then we must most certainly opt for Platonism instead of its rival: a host of sentences that are unambiguous within the Platonist framework are ambiguous on Moruzzi's view—e.g., sentences where musical works are said to be performed (cf. Moruzzi, 2018, p. 345); sentences where unperformed musical works are said to be composed (cf. Moruzzi, 2018, p. 346). In sum: if we are to abide by (Semantic Parsimony), we had better identify musical works with atemporal *abstracta*, and contend that some acts of artistic creation are acts of creative selection.

The type theorist's adherence to Lebens's ambiguity, too, is perfectly consistent with her endorsement of (Semantic Parsimony). As Lebens (2015 p. 3072) notes, terms of various kinds are patently ambiguous between processes and products. For example: when a dermatologist says that she is treating her patient's sunburn, 'sunburn' refers to a particular skin lesion; when a biophysicist said that she is investigating sunburn, 'sunburn' refers to the process whereby people acquire skin lesions after a lengthy exposure to ultraviolet light. Similarly, the term 'corrosion' refers on some of its occurrences to a product, and, on others, to a process.[57]

The type theorist believes that the properties that fictional entities possess (i.e., instantiates or kinstantiates) belong to four classes. First: they instantiate properties of the kind that Van Inwagen (2000, pp. 245–246) would call 'logical': they exist; they are self-identical, etc. Second: they instantiate properties merely in virtue of their ontological nature—they are abstract; they are universal; they are exemplifiable; etc. Third: they kinstantiate their fictional properties—by (Theorem 1). And fourth: they instantiate certain critical properties—e.g. having been creatively selected by an author, being realistic, etc. Nevertheless, it might seem as if none of this suffices to explain the truth of some intuitively true fictive sentences. Witness:

21.   Ireneo is the son of María Clementina Funes.
22.   Ireneo is younger than Nebúkadnessar Nebúkadnessarsson.

The type theory states that fictional entities are types. It suffices to inflate this state-

---

57. The postulation of this ambiguity between processes and products was originally made by Davies (2004, p. 187).

ment semantically to conclude that, on the type theorist's view, fictive names refer to types. If this conclusion is sound, then both 'Ireneo' and 'María Clementina Funes' in (21) refer to types, and so do 'Ireneo' and 'Nebúkadnessar Nebúkadnessarsson' in (22). If this is the case, then (21) states that Ireneo is the son of a type and, thus, the son of an abstract entity, whilst (22) says that Ireneo is younger than a type and, consequently, younger than a universal entity.

Clearly, the property of being the son of an *abstractum* is not a logical property, nor is the property of being younger than an *abstractum*. Moreover, neither of these properties is possessed by Ireneo merely in virtue of his ontological nature, and they are not critical properties. Finally, these properties do not belong to the set of Ireneo's fictional properties: evidently, Borges's protagonist is not portrayed, in the story where he appears, as possessing the queer quality of being the son of an abstract entity—nor, of course, the equally queer quality of being younger than an entity that is not located in spacetime. In sum: the properties in question do not belong to any one of the classes of properties that the type theorist attributes to Ireneo; and, therefore, it appears that we ought to conclude that the type theorist cannot vindicate our intuitions that (21) and (22) are true.

I submit, however, that this conclusion depends on a misunderstanding of (21) and (22). Consider:

23.   In Spanish, 'Q' is always succeeded by 'U'.
24.   The human dystrophin gene is longer than the U7 snRNA gene.

Clearly, there is no particular 'Q'-token that the term "'Q'" in (23) refers to, so it must refer to the 'Q'-type. By a parallel line of reasoning, we may wish to conclude that "'U'", too, refers to a type. If this is the case, then (23) says that, in Spanish, the 'Q'-type has the property of being always succeeded by the 'U'-type. Nevertheless, the 'U'-type is not located in spacetime, so it cannot succeed anything—neither spatially nor temporally. If (23) is true, as our intuitions say it is, then the property attributed to the 'Q'-type in (23) cannot be that of being succeeded by the 'U'-type.

In Spanish, every 'Q'-token x possesses the property of being such that there exists a y such that y is a 'U'-token and x is succeeded by y. Perhaps the property in question is shared; perhaps, that is to say, this property is instantiated by all 'Q'-tokens, and kinstantiated by the 'Q'-type. If this is the case, then (23) may be plausibly

113

read as saying that the 'Q'-type kinstantiates the property denoted by the following propositional function: "there exists a y such that y is a 'U'-token and x is succeeded by y".

It appears, in sum, that (23) is not to be read at face value: it appears to have the form of (25), but it is best understood as saying something along the lines of (26):

25. Type K_1 is related by relation R to type K_2.
26. Type K_1 kinstantiates the property denoted by the following propositional function: 'there exists a y such that y is a token of K_2 and x is related by R to y'.

Something similar ought to be said about (24). *Prima facie*, (24) contains two type-referring expressions: 'the human dystrophin gene' and 'the U7 snRNA gene'. If this is the case, then (24) says that the human-dystrophin-type is longer than the U7-snRNA-type. Nevertheless, the U7-snRNA-type is abstract; it does not have a length and, consequently, nothing is longer than it. If (24) is true, as our intuitions say it is, then the property attributed to the human-dystrophin-type in (24) cannot be that of being longer than the U7-snRNA-type.

Every human-dystrophin-token x possesses the property of being longer than every U7-snRNA-token. Perhaps the property in question is shared; perhaps, that is to say, this property is instantiated by all human-dystrophin-tokens, and kinstantiated by the human-dystrophin-type. If this is the case, then (24) may be plausibly read as saying that the human-dystrophin-type kinstantiates the property of being longer than every U7-snRNA-token. Like (23), then, (24) is not to be read at face-value: it appears to have the form of (25), but it is best understood as saying something along the lines of the following:

27. Type K_1 kinstantiates the property of being related by R to every token of K_2.

I think that (21) and (22) are just like (23) and (24). Let us see. (21), like (23), appears to have the form of (25), but it is best read, I contend, as having the form of (26). Under this reading, (21) says that Ireneo kinstantiates the property denoted by the following propositional function: 'there exists a y such that x is a María-Clementina-Funes-token and x is the son of y'. By (Axiom 1), María Clementina Funes is a type tokened by

Uruguayan women. Hence, the property attributed to Ireneo in (21) is that of being the son of a Uruguayan woman. That is, intuitively, one of Ireneo's fictional properties. By (Theorem 1), Ireneo (the type) possesses this property and, thus, (21) is true.

Now, (22), like (24), appears to have the form of (25); but I believe that it is best read as having the form of (27). Under this reading, (22) says that Ireneo kinstantiates the property of being younger than every Nebúkadnessar-token. In the story, Ireneo is portrayed as a nineteen-year-old man. Nebúkadnessar, by contrast, is portrayed as an older man. By (Axiom 1), Ireneo is a type tokened by nineteen-year-old men; and, by (Theorem 1), Ireneo himself (the type) possesses (kinstantiates) the property of being nineteen years old. By (Axiom 1), moreover, Nebúkadnessar is a type tokened by older men. Since Ireneo himself (the type) possesses the property of being nineteen, we may rationally attribute to it the property of being younger than every Nebúkadnessar-token. Thus (22), under the reading that I proposed for it, is true. (Attributing to a type the property of being younger than every Nebúkadnessar-token might appear rather odd, but I believe that it is not: recall that the attribution of a certain age to the type in question corresponds to the relation not of instantiation, but of kinstantiation. Moreover, we perform analogous attributions all the time: as I showed earlier, the human dystrophin gene (the type) is, in (24), attributed the property of being longer than every U7-snRNA-token.)

The type theorist, in sum, is capable of accounting for both (21) and (22) in a satisfactory manner: she need only read them in analogy with the way in which certain type-referring sentences (e.g. (23) and (24)) are best read. Importantly, the ambiguities that the type theorist appealed to in order to account for (21) and (22) are in line with (Semantic Parsimony) in that they both ought to be countenanced—and they ought to be countenanced, in particular, so that we can account for the truth of sentences like (23) and (24).

If the arguments that I have advanced in this section are sound, then the type theorist has attained three notable achievements. First: she has provided a satisfactory answer to the question of what entities are those which we represent in our minds when we imaginatively engage with a work of fiction. Second: she fulfilled both (Anti-Revisionism) and (Semantic Parsimony). And third: she accounted for the intuitive truth of tricky sentences like (21) and (22) without violating (Semantic Parsimony).

4   Final Arguments for the Type Theory

In Chapter Two, I argued that the creationist in incapable of construing certain fictive sentences as true. At this point, I believe, an objector to the type theory could contend that the type theorist struggles with exactly the same sentences. Let us see.

Halldór Laxness's *Lilja* is the story of a man called 'Nebúkadnessar Nebúkadnessarsson', and his friendship with a child called 'Lilja'. At some point, Nebúkadnessar says that Lilja is fond of raisins. Lilja's fondness of raisins is never mentioned again; and yet, in virtue of Nebúkadnessar's assertion, the following sentence seems to us to be true:

28.   Lilja is fond of raisins.

The creationist, we have seen, believes that fictional sentences are ambiguous. Thus she believes that (28) is ambiguous between (29) and (30):

29.   Lilja instantiates the property of being fond of raisins.
30.   Lilja holds the property of being fond of raisins.

Within the creationist's framework, (29) is untrue: Lilja, within this framework, is an abstract object, and abstract objects cannot instantiate the property of being fond of raisins. (30), by contrast, is true; and the reason why we perceive (28) as true, the creationist contends, is that we ordinarily use it to express (30) rather than (29). But is it really the case that (30) is true within the creationist's framework? I believe that it is not. The creationist adheres to the following:

Holding

A *fictum a* holds the property of being $G \leftrightarrow$ according to the relevant story, a is G.

According to (Holding), the *abstractum* Lilja holds the property of being fond of raisins just in case *it* is said in the story to be fond of raisins. In Chapter Two, I argued that it is not. In the story, Nebúkadnessar is pictured at some point as uttering 'Lilja is fond of raisins'. But what proposition is that which Nebúkadnessar expresses with his utterance? Not, I argued, the proposition that a certain *abstractum* is fond of raisins. As I noted earlier, abstract objects cannot be fond of raisins; and, in the scene where Nebúkadnessar's utterance is produced, Nebúkadnessar is not confused, nor is he speaking lies. Consequently, the proposition that he puts forward in this scene is not the proposition that a certain *abstractum* is fond of raisins. If this is the case, then the

abstractum Lilja is not said in the story to be fond of raisins. By (Holding), it follows that (28) is untrue.

The objector of the type theory could claim that the type theorist, too, struggles with (28). Her objection goes as follows:

> The type theorist believes that *ficta* are types, and, accordingly, that fictive names refer to types. On her view, then, 'Lilja' in (28) refers to a type. Nevertheless, when Nebúkadnessar uttered 'Lilja is fond of raisins', he was clearly not speaking of a type: he was speaking of a child! The question then arises of what, within the type theorist's framework, grounds the truth of (28).

This objection, to my mind, is genuinely interesting, for it reveals that (28) is a truly tricky sentence to account for. When Nebúkadnessar uttered 'Lilja is fond of raisins', it is unquestionable that he was speaking of nothing other than a child. This suggests that an ontologist of *ficta* may legitimately construe (28) as true just in case she construes 'Lilja' in (28) as referring to a child. Nevertheless, it is a fact that 'Lilja' in (28) refers to a fictional entity. This entity is determinately not an actual child: actual children are not fictional entities. But it seems that it is not a non-actual child either: in §3 (and more extensively in Chapter Three), I showed that Kripke's contention is decisive that *ficta* are not non-actual individuals. Trivially, if the referent of 'Lilja' in (28) is not an actual child, nor a non-actual child either, then it is not a child at all. And, if it is not a child, then it is wholly unclear why Nebúkadnessar's assertion that a certain child is fond of raisins would allow us to uphold (28).

If my contention above is sound, then (28) is, as I said earlier, genuinely tricky. It is so much so, I believe, that sentences like it deserve their own special attention. Unfortunately, I lack the space here to discuss them extensively. And yet, I wish to provide a clear indication that the type theorist, unlike the creationist, possesses proper resources to handle them. Suppose that I utter the following:

31.   At the end of a wedding ceremony, the priest says that the groom and the bride he declares to be husband and wife.

If I utter (31), I will be thought to have spoken truthfully. It is clear, though, that 'the priest' does not refer to a particular priest-token. It refers, I submit, to a type. In accordance with Wollheim's contentions, moreover, we think of this priest-type as if it were an archetypical priest-token when we read (31). Now, 'the groom' and 'the

bride', too, appear to refer to types rather than tokens. And yet, it is clear that (31) does not depict our archetypical priest as performing the downright oddity of declaring two universal *abstracta* to be husband and wife. The property that is attributed to our priest in (31) is, of course, the property of having said, of a groom-token and a bride-token, that he declares them to be husband and wife.

Now, on the type theorist's view, *ficta* are types, so Nebúkadnessar and Lilja are types. By (Axiom 1) Nebúkadnessar is, in particular, a type instantiated by non-actual Icelandic men; and Lilja, a type instantiated by non-actual Icelandic children. Now consider:

32.   Nebúkadnessar says that Lilja is fond of raisins.

(32), I contend, is to be read in a way exactly analogous to the way in which we read (31). On this reading, the property attributed to Nebúkadnessar in (32) is not that of having said of the Lilja-type that it is fond of raisins: it is the property of having said of a Lilja-token (i.e., a child) that it is fond of raisins. This is exactly analogous to the fact that the priest in (31) is attributed the property of having wedded not two types, but a groom-token and a bride-token.[58]

The type-theorist, in sum, is able to construe our *Lilja* scene in a way that honours our intuitions that Nebúkadnessar was speaking of a child: she need only interpret (32) in line with the way in which we naturally interpret structurally analogous pieces of type-discourse—e.g., (31). But can the type theorist interpret (32) in such a way that it gives us the grounds to uphold (28)? I believe that she does.

Intuitively, our commitment to the truth of (31) sanctions us to conclude (33):

33.   At the end of the wedding ceremony, the groom becomes husband.

As I noted earlier, (31), in the surface, appears to concern three types. Thus it appears to have the following form:

34.   At t, K_1 performs a speech act whereby K_2 and K_3 are attributed properties G_1 and G_2 respectively.

---

58. My contention here relates, of course, to the contention that I advanced in the previous section that some of the sentences that appear to have the form of (25) are best understood as having the form of (26).

I implied earlier that (31) does not really have the form of (34), for the priest's speech act does not concern types. And yet, (31) allows us to derive (33), which has the following form:

35. At t, K_2 is G_1.

Now, on the type theorist's view, *ficta* are types. For her, then, (32), in the surface, appears to have the following form:

36. K_1 performs a speech act whereby K_2 is attributed property G.

Earlier, I implied that (32) does not really have the form of (36), for I said that Nebúkadnessar speaks not of a type, but of a child. And yet, I believe, our commitment to (32) allows us to derive (28). (28), for the type theorist, has the following form:

37. K_2 is G.

It is evident that (36) and (37) relate logically to each other in exactly the same way as (34) and (35). Consequently, it appears, our inference from (32) to (28) may well parallel that from (31) to (33). (The only difference between these inferences appears to be the following: Nebúkadnessar's speech act is constatative: he *asserts* that Lilja is fond of raisins. Our knowledge that this assertion is truthful allows us to conclude (28). The priest's speech act, by contrast, is performative: he *declares* the groom and the bride to be husband and wife. Our knowledge that he has the relevant authority to do so allows us to conclude that (33) is true.) If the inference from (32) to (28) does indeed parallel that from (31) to (33), then the former is as legitimate as the latter.

Of course, what exactly the semantics is that which lies behind the inferences in question is, at the moment, rather unclear. Describing this semantics, moreover, requires substantial amounts of work that I, unfortunately, do not have the space to carry out here. Despite this, we have weighty reasons to believe that, within the type theorist's framework, the inference from (32) to (28) is as legitimate as an obviously legitimate inference—viz. (31) to (33): both of these pairs of sentences are, on the surface, analogous to one another—as shown by (34)-(37). Hence, whatever the semantics is that which legitimises one of the corresponding inferences may well be used to legitimise the other.

If my arguments above are sound, then treating (32) as we naturally treat structurally alike pieces of type-discourse (e.g., (31)) will plausibly lead the type theorist to legitimise our commitment to (28). Since (28) is as tricky as I earlier proved it to be, the type's theorist's ability to handle it is, I submit, a truly remarkable feat.

Before concluding, I would like to point to a number of directions in which the type theory could be exploited in order to solve several of the very many philosophical puzzles that fiction, and artistic creation more generally, produce.

To some ontologists, fictional entities seem to be ontologically incomplete: there are determinables such that *ficta* possess them, but *ficta* do not possess any determinate of those determinables. Consider, for instance, the following determinable: x has wisdom teeth or x does not have wisdom teeth. According to the ontologists that I am referring to, Ireneo bears this determinable, like any flesh-and-blood human does, but he does not bear any determinate corresponding to it: there is simply no matter of fact as to whether he has wisdom teeth or not. Types are exactly like this. The tiger-type, for instance, possesses the following determinable: x is a male or x is a female—just like each one of its tokens. And yet, it does not possess any determinate corresponding to this determinable—as demonstrated by the fact that both (38) and (39) are false:

38.   The tiger is a large female felid.
39.   The tiger is a large male felid.

Identifying *ficta* with types, then, may help the ontologists in question account for their incompleteness.

Now, one of the most powerful threats ever faced by fictional realists is constituted by Everett's putative *reductio ad absurdum.* In his (2020b), Friedell, on the grounds of Woodward's work (2017), suggested that the fictional realist may escape Everett's *reductio* if she identifies *ficta* with abstract entities existing atemporally and platitudinously. That is exactly what the type theorist does.[59] I believe, therefore, on the basis of Friedell's work, that a full-fledged proof that the fictional realist can indeed overcome Everett's threat can be construed on the basis of the type theory.

And finally: nowadays, the most widely-held view concerning the ontology of music

---

59. The existence conditions that I attributed to types in §1 of Chapter 4 evidently imply that types exist platitudinously.

states that musical works are types. On the basis of this, and the arguments I advanced above in favour of the type theory of *ficta,* I believe that the project is feasible of constructing a uniform ontology for all kinds of abstract artistic creations: not only *ficta* and musical works, but also literary works, drama, etc.[60] In the following chapter, I will take one substantial step towards the completion of this project.

5   Conclusion

The puzzles of fiction are so many that there is not enough space in one chapter to discuss them all. If my arguments above are sound, however, then the type theory is promising. It is so much so, I believe, that it is worth exploring whether it might foster novel and interesting solutions to the puzzles that I have not had the space to address in this chapter.

As I contended at the end of the previous section, the project appears feasible of assigning all sorts of abstract artistic creations to the category of types. In the following chapter, I will take one substantial step towards the completion of this project. In doing so, I hope to show that the type theory of *ficta* is very much apt to fit into a promising project of construing a uniform ontology for all sorts of abstract artistic creations. I aim to thereby provide one further motivation for the theory in question.

---

60. For an argument to the effect that these artistic creations are abstract, see the introduction of Chapter Six.

Chapter Six:
Musical Ontology and the
Audibility of Musical Works

Some artistic creations are concrete: sculptures and paintings, for instance, are identical to spatiotemporally located objects. Literary works and performing artworks, by contrast, appear to be abstract instead, for there are no concrete objects that they can be identified with. Consider, for example, Rulfo's *Pedro Páramo* and Beethoven's *Waldstein* sonata. *Pedro Páramo* cannot be identified with any one of its copies: its copies can be burned, torn up, or otherwise physically destroyed, but *Pedro Páramo* cannot. *Waldstein* cannot be identified with any one of its scores for the same reasons—but it cannot be identified with any one of its performances either. *Waldstein* is repeatable: it admits of multiple performances and, hence, multiple occurrences. Its performances, on the other hand, do not. Arrau's performance of *Waldstein* at the Beethovenfest in Bonn in 1971, for example, could be imitated, and recordings of it can be reproduced; but it happened at a certain place, at a certain time, and it cannot happen again.

In an ideal scenario, a theory exists which assigns all of these abstract artistic creations to one and the same ontological category: such a theory would be insuperably parsimonious. And, to my mind, there is an ontological category which appears particularly suitable for the purpose of articulating the theory in question: the category of types.

Types are abstract, universal entities whose instances are called 'tokens'. They are different from other universal entities because they are related to their instances in a special way (see §1 in Chapter Four for an extensive characterisation of types). As I noted earlier, musical works are repeatable. In his (1980), Wollheim famously formulated an argument to the effect that, amongst the various sorts of repeatable entities

that exist, only types relate to their occurrences in the same way in which musical works relate to theirs. Nowadays, the most widely-held view concerning the ontology of musical works states that musical works are types of concrete sound events—types, that is to say, tokened by performances. Kivy (1983, 1987, 2002, pp. 212–223) endorsed this view after Wollheim; and, afterwards, multiple authors have developed sophisticated versions thereof (see, e.g., Levinson, 1980; Wolterstorff, 1980; Howell, 2002; Dodd, 2000; 2002; 2004; 2007; 2008; Walters, 2013; Puy, 2019; 2021). Now, artistic creations other than musical works have been identified with types, too: literary works, for instance, have been claimed to be types of sentence-token sequences (see, e.g, Wollheim, 1980, pp. 49–50); and, in Chapter Five, I defended a type theory of *ficta*. In sum, then, there are abstract artistic creations of various kinds which may be accounted for, with plausibility, in terms of types. If this is the case, then perhaps types may be used to account for *all* abstract artistic creations.

A crucial step towards the aim of construing all of the creations in question as types consists in using the framework of the musical type theory to account for all performing arts. And, to my mind, there are *prima facie* persuasive reasons to believe that this step is attainable. Trivially, the works that belong to the category of the performing arts are all characterised by being capable of having performances. If the type theory accounts adequately for the relation between musical works and their performances, then it may well account satisfactorily for the relation between the other performing arts and *their* performances. Nevertheless, the musical type theory faces two substantial problems, and one may sensibly expect the type-theoretic accounts of the other performing arts to face analogous difficulties. Thus, if we want to successfully use the framework of the musical type theory to account for all of the performing arts, we ought to resolve the problems in question.

The first of these problems is the creatability problem. Types exist outside of spacetime. If a musical work is a type, then there does not exist a point in time where it was brought into existence. If this is the case, then why are we convinced that musical works are created by composers at specific points in time? This question constitutes the creatability problem.

The creatability problem has been convincingly addressed by Kivy (1983; 1987) and Dodd (2007, pp. 99–142), and the problem that I am presently interested in, accordingly, is the other one: the audibility problem. The musical type theorist claims that musical

works are types and, hence, that musical works are abstract. Nevertheless, the thesis that musical works are abstract clashes with the seemingly platitudinous thought that these works are ordinarily appreciated by means of audition. The things that stimulate our auditory organs are sounds—viz. spatiotemporally located physical phenomena that are causally related to our perceptions. *Abstracta*, by contrast, are orthodoxically thought to be causally inert entities that are not located in spacetime, so they appear to be not at all like the objects that produce auditory experiences. This gives rise to the audibility problem: how can musical works be appreciated by means of audition if they are abstract? This chapter I will devote to proposing a novel answer to this question.[61]

Dodd (2007, pp. 11-16, 92-94) and all the other musical type theorists (e.g., Puy, 2019, p. 249) have employed one and the same strategy to respond to the audibility problem: they have defended the claim that musical works are appreciated by means of audition because they are audible despite being abstract. In §1 and §2, I will argue that this strategy fails—and I will show that it does so even in a perfectly ideal formulation—: musical works, I will contend, simply cannot be construed as audible entities. Then, in §3, I will advance a novel response to the audibility problem: I will show that musical works are appreciated by means of audition despite the fact that they are not audible.[62] If my arguments in this chapter are sound, then the musical type theorist has satisfactory solutions to both the creatability problem and the audibility problem; and, hence, the framework of her theory may be safely used to account for performative artworks in general.

1 Deferred Audition

The most thorough of Dodd's defences of the musical type theory (2007) comprises what is probably the most elaborate solution to the audibility problem that has ever been proposed (2007, pp. 11–16, 92–94). This chapter is devoted to it.

According to Dodd, "it is surely a datum that musical works are things that we can listen to" (2007, p. 12). And, "[w]hen listening to a performance of a work of music,

---

61. §1–§3 of this chapter constitute a paper that has been published in the *British Journal of Aesthetics* (cf. Meléndez Gutiérrez, 2022).

62. Not all musical ontologists believe that musical works are abstract: some think that they are concrete objects like performances (see, e.g., Moruzzi, 2018) and fusions of performances (see, e.g., Caplan and Matheson, 2006). I will not discuss their proposals here, however, for my aim in this paper is to address the problem of how musical works are appreciated by means of audition *if they are abstract*.

one thereby listens to the work performed". He follows earlier musical type theorists in making this claim: according to Levinson, "musical works [...] *can* be heard in or through their performances" (1980, p. 27, emphasis in the original). Wolterstorff, similarly, states that, "in listening to a symphony one hears two things at once, the symphony and the performance thereof" (1980, pp. 40–41). How we hear a musical work by hearing one of its performances, Dodd maintains, is easily explained by analogy with Quine's deferred ostension (1969, pp. 39–41). According to Quine, it is possible to refer to a type by demonstrating one of its tokens: one can, for instance, point to a letter-token alpha, say 'this is alpha', and successfully refer to the relevant letter-*type*. "Well", Dodd says, "as for demonstrative reference, so for perceptibility" (2007, p. 12): just as we can refer to a letter-type by demonstrating one of its letter-tokens, Dodd says, by hearing a performance of a musical work we can hear also the type that stands behind it (2007, p. 12).

Dodd, then, claims that audition must be capable of reaching the type that *stands behind* a performance, just as reference is capable of reaching the type that stands behind a letter-token.[63] To my mind, though, this claim is nothing more than a stipulation: Dodd gives us no reason to believe that what is true of reference must be true also of perception, and it certainly does not appear to be so: it is clear, for instance, that some things can be referred to which cannot be perceived—e.g. the empty set.[64]

More importantly, Dodd's analogy with deferred ostension fails to address the difficulties of metaphysical order that are involved in the hypothesis that musical works are audible. For one thing, the orthodoxy has it that abstract objects are causally inert (see, e.g. Dummett, 1973, p. 493; Bach, 1987, p. 12; Linsky & Zalta, 1995, p. 252; Balaguer, 1998, p. 1; Van Inwagen, 2007, p. 200; Parsons, 2008, p. 1). If *abstracta* are indeed causally inert, then musical works cannot cause auditory experiences.

Dodd discusses this issue shortly after presenting his Quinean analogy (2007, pp. 13–16). He abides by the orthodoxy in stating that the *relata* of causal relations are events, not objects—neither abstract nor concrete (see, e.g., Davidson, 1963; 1967). Nevertheless, following Bennett (1988, p. 23), and Burgess and Rosen (1997, p. 24), Dodd claims that

---

63. In a recent paper (forthcoming), Cohen discusses Dodd's analogy, develops it in a very interesting direction, and argues that, thus developed, the theory has at least two problematic consequences.

64. Even Maddy, who believes that some sets are perceivable, admits that some others, like the empty set, are not (Maddy, 1990, pp. 156–157).

objects can be correctly said to be causally efficacious in a number of circumstances nonetheless. After discussing what these circumstances might be, he concludes, with Caplan and Matheson (2004, p. 121), that "[...] all we can say is that an object so participates causally in an event by being *appropriately related* to said event" (2007, p. 15, emphasis in the original). Naturally, Dodd says, just when an object holds an appropriate relation to an event in the relevant sense is up for grabs; but, plausibly, such an appropriate relation obtains when the object in question is a type, and the event is one of its tokens (2007, p. 15). If this is correct, then a musical work may be said to exercise causal powers when one of its token sound events exercises causal powers. It follows that, if a sound event causes an auditory experience, then the musical work that this sound event is a performance of, too, can be said to be the cause of the auditory experience in question.

Dodd adduces ordinary language data in order to support his contention (2007, pp. 15–16). A film, he maintains, is a type, and its tokens are showing-events. Now suppose that the showing of a film sparked a riot. We have seen that Dodd contends that a type may be said to exercise causal powers where one of its tokens does. If this is right, then, in the hypothetical scenario he asked us to imagine, not only the showing, but the film itself may be said to have caused the riot. Ordinary language ratifies this result: it would be entirely natural, Dodd notes, to claim that the riot was caused by the film.

Juvshik (2018), though, thinks that Dodd's contention must not be granted. He agrees with Dodd that ordinary language allows us to speak of types as causally efficacious: it allows us to speak, for instance, of films causing riots. "[B]ut from the fact that we talk a certain way", Juvshik says, "it does not follow that things are that way" (2018, p. 812). According to Juvshik, there are good reasons to reject the view that musical works have causal powers: drawing from Kim's exclusion arguments against mental causation (1993; 1998), he advances an argument from parsimony that states, briefly, that the auditory experiences that we have when we listen to music can be exhaustively explained in terms of the physiological structure of the human auditory capacity, and the behaviour of the sound waves that constitute performances (2018, pp. 814–816). If this is the case, then there is no explicative work left for musical works to do; and, for the sake of parsimony, we should refrain from attributing causal powers to them.

In his (2020a), Friedell tried to vindicate Dodd's contention; but I do not think he succeeds. Friedell maintains, *contra* Juvshik, that there *is* a reason to attribute causal

126

powers to musical works: namely, it is intuitively the case that types are causally effi-cacious. Harriet Beecher Stowe's novel *Uncle Tom's Cabin*, Friedell says, is an instance of this: it is a type, for it is not located in spacetime, and it has multiple occurrences; and yet, it is intuitively true that it caused many Americans to support the abolitionist cause. Juvshik, of course, would insist that the fact that this seems intuitively true does not mean that it is true; but Friedell has more to say. Perhaps, as Juvshik would have it, *Uncle Tom's Cabin* did not cause many Americans to support the abolitionist cause; but, as a matter of fact, many Americans supported the abolitionist cause. If *Uncle Tom's Cabin* did not cause this, what did? The *copies* of *Uncle Tom's Cabin*, perhaps. Perhaps, that is to say, utterances of (1) should be read not at face value, but as metonymical for (2):

1. *Uncle Tom's Cabin* caused many Americans to support the abolitionist cause.
2. The copies of *Uncle Tom's Cabin* caused many Americans to support the abolitionist cause.

Unfortunately, Friedell says, (1) fails two simple tests for metonymy. Consider:

3. The sax has the flu.

It seems clear that the expression 'The sax' in (3) is to be read not literally, but metonymically: it refers to a saxophonist, not to a sax. (3), accordingly, passes both of Friedell's tests for metonymy: it is felicitous to attribute properties to the referent of 'the sax' under the assumption that 'the sax' is to be read as metonymical; it is infelicitous to attribute properties to it under the assumption that 'the sax' is to be read literally. Witness:

4. The sax has the flu; she will be back next week.
5. The sax has the flu; it is covered in bacteria.

(4) could only be asserted under the assumption that 'the sax' is metonymical for a saxophonist—cf. the pronoun—, and it is felicitous. (5), on the other hand, could only be asserted under the assumption that 'the sax' refers to a sax—cf., again, the pronoun—, and it is infelicitous. Hence, (3) passes both metonymy tests, just as we would expect. (1), by contrast, passes neither:

7. *Uncle Tom's Cabin* caused many Americans to support the abolitionist

cause; many of them have been lost and will never be found.

8.   *Uncle Tom's Cabin* caused many Americans to support the abolitionist cause; it was the most popular novel of the nineteenth century.

It is infelicitous to attribute properties to the referent of '*Uncle Tom's Cabin*' under the assumption that '*Uncle Tom's Cabin*' refers metonymically to the copies; it is felicitous to attribute properties to it under the assumption that '*Uncle Tom's Cabin*' refers to the type. Thus, Friedell concludes, we must conclude that (1) is not metonymical, but literally true, and types, therefore, are causally efficacious.

I believe, however, that both of Friedell's tests are flawed. Consider:

8.   Mozart has charmed listeners for well over two centuries.

(8) is evidently metonymical: Mozart, the man, might have been charming, but he was only thirty-five when he died, so he surely did not charm people for more than two hundred years. It is clear, then, that 'Mozart' in (8) refers metonymically to Mozart's compositions—and yet, the following is infelicitous:

9.   Mozart has charmed listeners for well over two centuries. They include over fifty symphonies, and twenty-seven piano concertos.

There are, then, *pace* Friedell, uncontentiously metonymical expressions whose referents cannot be felicitously attributed properties under the assumption that they are indeed metonymical. Now consider (10):

10.   I was reading Alfonso Reyes this morning.

(10), too, is undoubtedly metonymical: the occurrence of 'Alfonso Reyes' in this sentence refers not to the man, but to his works. If Friedell's second test were accurate, we could expect it to be infelicitous to attribute a property to the referent of 'Alfonso Reyes' under the assumption that 'Alfonso Reyes' is to be read as referring to the man; but then Friedell's second test predicts the wrong result, for (11) is entirely felicitous:

11.   I was reading Alfonso Reyes this morning. He is becoming one of my favourite essayists.

If my arguments are sound, neither of Friedell's tests can be trusted to tell metonymical expressions from literal ones. The fact that (1) fails them both, then, is not a reliable

indication that (1) is not metonymical. Friedell has thus given us no reason to suppose that (1) is literally true, nor, therefore, that types have causal powers. But then Juvshik is right: given that our auditory experiences can be exhaustively accounted for by attributing causal powers to performances only, there is no reason to suppose that musical works themselves are causally efficacious. By the principle of parsimony, we should refrain from doing so.[65]

As Davies (2009, p. 101) maintains, however, the main metaphysical obstacle faced by Dodd in his project of explaining how musical works *qua* types are audible is not the doctrine that *abstracta* are causally inert, but the fact that, according to Dodd's own construal of types, these entities do not bear *acoustic properties*—i.e., properties like being in a certain key, or featuring a certain chord or rhythmic figure.

Like Wolterstorff (1980, p. 61), Dodd adheres to the doctrine of analogical predication, which states that predicates denoting acoustic properties can never be applied univocally to both musical works and their performances (2007, pp. 83–85). Schematically, this doctrine has it that, if the predicate 'is F' denotes an acoustic property, and it may be correctly applied to both a musical work W and any one of its correct performances, then 'is F' is applied analogically: the property attributed to performances of W by means of an application of 'is F' is the acoustic property of being F; but the property attributed to W itself is the *normative* property of being such that nothing can be a correct performance of W if it is not F. The predicate 'begins with a C-sharp minor' chord, for instance, may be correctly applied to Sibelius's *Finlandia* and any one of its correct performances; but the properties thereby attributed to them are distinct: the property attributed to performances is the acoustic one of beginning with a C-sharp minor chord; the property attributed to *Finlandia* is its normative counterpart: being such that nothing can be a correct performance of *Finlandia* if it does not start with a C-sharp minor chord.

The motivation behind this doctrine is the conviction that types, and thus musical works, are unstructured; the upshot is that musical works do not bear acoustic properties at all, but only the associated normative ones. They are, then, not types *simpliciter*, but *norm*-types (cf. Wolterstorff, 1980, p. 56). If this is the case, though, then the following question arises: how do we listen to musical works if they possess no acoustic

---

65. For further literature on the question whether there are causally efficacious *abstracta*, see Walters (2013); Juvshik (2020); Aliyev (2020; forthcoming).

properties?

Dodd responds by saying that our locutions concerning audition are ambiguous (2007, pp. 92–94). Acoustic properties, on Dodd's view, are possessed by sound events, which are *directly* audible; but several *bona fide* objects of audition are not sound events: birds, bells and musical works, for instance, do not belong to this ontological category, so there must be a second sense in which things may be said to be audible.

Dodd, we have seen, believes that not only events, but also objects, may be rightly said to be causally efficacious: an object, he says, counts as causally efficacious when it bears an *appropriate relation* to an event that causes an effect. Later on, Dodd claims that, when an object holds this kind of relation to a *sound* event, this object may be said to be audible in a *derivative* sense. As I said earlier, Dodd thinks that the relevant relation holds between an object and an event when the object is a type, and the event is one of its tokens. Performances are sound events, and they are tokens of musical works; so, when a musical work W is performed, a sound event comes into existence that W is appropriately related to and, hence, W can be said to be derivatively audible. On Dodd's view, then, musical works are audible, but not in the same sense as performances. According to Dodd, however, this is unimportant: the sense in which musical works are audible just is the sense in which many familiar objects of audition are audible. As I noted earlier, birds and bells, for instance, are not sound events and, consequently, not directly audible, but audible only in the sense that they are appropriately related to tweeting and chiming events, respectively.

A well-known test for ambiguity, devised by Zwicky and Sadock (1975), tells against Dodd's contention that our locutions concerning audition are ambiguous. When one single occurrence of a polysemic expression has multiple senses, we encounter a zeugma. Consider:

12. The colour and the feathers were light.

In (12), one single occurrence of 'light' has two distinct senses, and the fact that (12) is, thus, a zeugma strikes us as obvious: there is something odd about it cannot go unnoticed. Now contrast (12) with (13):

13. I spent the morning listening to Bach's Brandenburg Concertos and a bird tweeting.

Tweeting events are sound events and, hence, according to Dodd's theory, directly audible; musical works, on the other hand, are only derivatively so. It follows that, on his view, the one occurrence of the verb 'to hear' in (13) has two different senses, corresponding to direct and derivative audition respectively. If Dodd were right, we could expect (13) to have a zeugmatic character; but it does not: there is nothing odd about it, nor anything whatsoever that suggests that the one occurrence of the verb 'to hear' that figures in it does indeed possess various senses.

Whether this test for ambiguity is accurate is a controversial matter. Predelli (2011, p. 276) and Ostertag (2012, p. 366) appealed to it to try to disprove the ambiguities posited by the doctrine of analogical predication; but Dodd and Letts (2017, pp. 252–253) argue that they both fail. In order to do so, they adduce Sennet's claim (2021) that some ambiguities might be too subtle for the folk to perceive them. On his view, when one single occurrence of a very subtly ambiguous expression has multiple senses, the sentences where it figures in a zeugmatic position do not acquire a zeugmatic tone: they do not appear odd, nor striking in any way. Sennet adduced the following example to substantiate his case:

14.   Toronto and numbers exist.

(14), Sennet notes, does not appear zeugmatic to ordinary speakers of English; but ontological pluralists maintain that the one occurrence of the verb 'to exist' in (14) has two distinct senses. Some ambiguous expressions, he concludes, will not pass Zwicky's and Sadock's ambiguity test.

Nevertheless, Ostertag (2012, pp. 372–373) rightly points out that ontological pluralism is highly contentious and, therefore, there is no certainty whatsoever that the verb 'to exist' is polysemous. The reason why (14) does not appear zeugmatic, then, may well be that it is not zeugmatic. If this is the case, then there is no reason to accept Sennet's claim that Zwicky's and Sadock's ambiguity test is unreliable. The truth of this claim, Ostertag contends, cannot be granted until an uncontroversial ambiguity is found which fails to test positive.

Dodd and Letts (2017, p. 243) think that Ostertag's position is problematic. He maintains, we have seen, that the reliability of Zwicky's and Sadock's ambiguity test cannot be challenged on the basis of philosophical convictions like ontological pluralism. Thus, Ostertag refuses to contemplate the possibility that philosophical inquiry could uncover

ambiguities not uncovered by this test. According to Dodd and Letts, Ostertag's position embodies, therefore, the very dubious doctrine of *linguistic deferentialism*, which states linguistic claims have a decisive bearing on philosophical disputes, and are, by default, better justified than philosophical hypotheses.[66]

Afterwards, Dodd and Letts (2017, p. 253) admit that a failure in Zwicky's and Sadock's test provides "some evidence" against the existence of the relevant ambiguity; and suggest that such evidence may be superseded by philosophical considerations. On pain of inconsistency, then, Dodd would have to admit that the fact that (14) fails the test tells against his contention that our locutions concerning audibility are ambiguous. He could insist that this result may be outweighed by philosophical considerations; but I do not think that there are any such considerations for Dodd to adduce.

Kripke (1977, pp. 268–269) proposes a test for accepting the existence of an ambiguity that differs from Zwicky's and Sadock's. He notes that positing ambiguities allows escaping theoretical trouble illegitimately, for any putative counterexample to any hypothesis may be dismissed by claiming that some key term is used in a special sense, different from its use in the relevant hypothesis. Consequently, Kripke maintains, we should counsel a policy of caution: do not posit an ambiguity unless you are truly forced to (1977, p. 269). In sum: in order to reduce the risk of incurring illegitimacy to the bare minimum, an expression may be counted as ambiguous just in case we are inescapably compelled to maintain that it is ambiguous. This claim seems to me to be solidly grounded in the principles that regulate theoretical inquiry: it amounts to the idea that senses ought not be multiplied beyond necessity; that they should be multiplied only when there is a compelling reason to do so. Kripke's second test, therefore, is founded on the principle of parsimony and the principle of sufficient reason, and appears, therefore, to be highly plausible. Let us, therefore, investigate what it can tell us about Dodd's ambiguity theory.

If hearing consists in perceiving the acoustic properties of an entity with our auditory organs, then the question arises of why we often speak of hearing things that undoubtedly lack acoustic properties, like birds and bells. This question may be answered in Dodd's way: maybe there is another sense in which we hear things. But the alternative is also available to us to maintain that we never *really* claim to listen to objects that lack acoustic properties; that, typically, our claims about listening to these objects are to

---

66. Daly and Liggins (2011) discuss and criticise this doctrine.

be understood as metonymical assertions that things are heard that *do* possess acoustic properties—viz., sound events. On this view, ordinary utterances of (15), for instance, express (16):

15.　　I heard bells.
16.　　I heard bells chiming.

Importantly, this metonymy theory is independently motivated: it seems that a speaker in their right mind would never assert (15) and deny that they meant to express (16). The metonymy theory, then, is not only adequate to account for the datum that we often speak about listening to objects that lack acoustic properties: it is also supported by our intuitions.

If the metonymy theory is available, then Dodd is not forced to commit to the hypothesis that our locutions concerning audition are ambiguous. These locutions, therefore, fail Kripke's second test for ambiguity. Dodd could, of course, simply refuse to comply with the results of this test; but, given that it is grounded on the principles of parsimony and sufficient reason, such a move would be rather reckless. Importantly, moreover, it is unclear why it should persuade *us* to adhere to Dodd's ambiguity theory: even if Dodd refuses the results of Kripke's test, the fact is that he has presented us with no motive to prefer his theory over the metonymical alternative. If this is the case, and I am right that the metonymy theory is independently motivated, then it is clear what the theoretical balance tips in favour of.

I have shown that Dodd's project of establishing that musical works *qua* types are audible faces two obstacles: the doctrine that *abstracta* are causally inert, and the fact that types, according to Dodd's own construal thereof, bear no acoustic properties. I argued that there is no easy way for him to overcome either of these obstacles: in order to overcome the first, he must uphold the wholly unmotivated hypothesis that musical works are causally related to our auditory experiences; in order to overcome the second, he must opt for his ambiguity theory instead of a theory that, at this point, appears visibly superior. Overcoming the obstacles in question, in sum, comes for Dodd at a very high methodological cost. In any case, moreover, it is unclear why *we* should be persuaded to follow him.

2   The ideal ontology

In this section, I will discuss a hypothetical ideal theory of the ontology of musical works. This theory, like Dodd's, identifies musical works with a peculiar kind of type that is tokened by sound events; but, by stipulation, its advocates face neither of the obstacles faced by Dodd in his attempt to establish that musical works are audible: the entities that musical works are identified with on this ideal theory, despite belonging to the realm of the abstract, have been proved to possess causal powers and acoustic properties. In particular, a musical work, according to this theory, bears all and only those acoustic properties that must be borne by any one of its correct performances. In what follows, I will contend that, even if musical works are identified with entities of this ideal sort, the hypothesis that they are audible will not go through.

It is undoubtedly hard to imagine how types of any kind could bear acoustic properties and, perhaps there are no such entities at all; but let us, for the sake of the argument, suppose that they exist. Beethoven's *Pathétique* sonata begins with a C-minor chord. On the ideal theory, it *really* begins with a C-minor chord. All correct performances of the *Pathétique*, too, begin with a C-minor chord; but there is an important disanalogy between the initial chord of the sonata and that of its performances.

Clearly, the initial chord of Brendel's 1975 performance of Beethoven's *Pathétique* has a maximally determinate duration. The initial chord of the sonata itself, by contrast, does not: there is no real number x such that the initial chord of the sonata lasts x seconds. On our ideal theory, musical works bear all and only those acoustic properties that must be borne by each one of their correct performances. Hence, if there were a real number x such that the initial chord of the *Pathétique* sonata lasted x seconds, the initial chord of each of its correct performances would last x seconds; but that is not the case: the initial chord of Brendel's performance and that of Horowitz's 1963 performance have patently different durations, and neither of them is incorrect.

The initial chord of Brendel's performance has a maximally determinate pitch. The initial chord of the *Pathétique* itself, by contrast, does not. If it did, then the pitch of the initial chord of each one of the correct performances of this sonata would be the same as the pitch of all the others; but it is not: the *Pathétique* sonata may be correctly performed, roughly, in A440, A441, A442, A443 and A444. Similarly, the initial chord of Brendel's performance has a maximally determinate volume, but the initial chord of

the sonata itself does not: if it did, then the initial chord of each one of the correct performances of the sonata would have exactly the same volume as all the others—but that is not the case: the initial chord of Brendel's performance, and that of Ashkenazy's 1997 performance have noticeably different volumes, and neither of them is incorrect.

There is nothing special about the initial chord of Beethoven's *Pathétique*, though. Beethoven's authoritative score of this sonata features a *tempo* indication for its first movement: '*Grave*'. It also states unequivocally that the initial chord of the sonata is a *fortepiano* C-minor chord. And yet, we have seen, the initial chords of its various correct performances vary somewhat with regard to duration, pitch and volume; so the initial chord of the *Pathétique* itself has no maximally determinate volume, pitch or duration. Moreover, the remainder of the authoritative score of this sonata is not any more specific in its remarks, nor are the scores of most other musical works: ordinarily, scores do not constrain the interpretative freedom of performers more than does the score of Beethoven's *Pathétique*. Furthermore, interpretative conventions, like scores, do not, or at least not usually, determine that all of the correct performances of a given musical work must be characterised by the very same durations, pitches or volumes. Hence, musical works, in very general terms, do not possess maximally determinate volumes, pitches or durations.

This is problematic for those who wish to maintain that musical works are audible. The things that stimulate our auditory organs are sounds; and every single sound, in virtue of its mere physical constitution, possesses maximally determinate acoustic properties. Every time we hear a sound, it has a maximally determinate duration, a maximally determinate pitch, and a maximally determinate volume—even if we cannot tell what exactly they are. If this is the case, however, then musical works, which possess no maximally determinate acoustic properties, are not the kind of thing that can be perceived with our auditory organs. The first chord of Beethoven's *Pathétique*, for instance, is a *fortepiano* C-minor chord that lasts the equivalent of five semiquavers and a half; but it has no maximally determinate volume, pitch, nor duration; and there is no such thing as listening to a sound with no maximally determinate duration, pitch or volume.

There are two independent replies that the advocates of the audibility of musical works could advance at this point. The first is that I am wrong to maintain that musical works lack maximally determinate acoustic properties. I derived this claim from two premises:

(i) the various correct performances of a musical work may vary somewhat with regard to duration, pitch and volume; and (ii) a musical work bears all and only those properties that must be borne by all of its correct performances. (i) states merely an easily empirically verifiable fact; but perhaps (ii) is false and, hence, my argument for the conclusion that musical works lack determinate acoustic properties is unsound. Whether (ii) is false or not, though, is, in reality, unimportant: ultimately, the burden is on whoever maintains that musical works have maximally determinate acoustic properties to provide a principled response to the question of what these properties are: what is the maximally determinate duration of the initial chord of Beethoven's *Pathétique*, for instance? What is its maximally determinate volume? I suspect that none of these questions can be answered with plausibility.

The second possible rejoinder to my argument says that musical works are *audible*, in some important sense at least, even if they lack maximally determinate acoustic properties. The picture that I have been sketching of the properties that characterise a musical work W and those that must be possessed by any one of its correct performances is the following: the properties that must be possessed by any one of the correct performances of W are *determinates of the determinables* that characterise W. My opponent, then, could say something along the following lines: plausibly, perception of a determinate *reaches through* to its determinable: after all, it would be severely counterintuitive to maintain that, when one hears the C-minor chord with which Brendel starts his performance of the *Pathétique*, with its maximally determinate pitch, one cannot rightly claim to have heard a C-minor chord *simpliciter*. If this is right, then, when we hear the determinates that characterise a performance of W, audition reaches through to the determinables that characterise W.

If this is correct, then it seems that Dodd was on the right track when he proposed that perception functions very much like Quine's deferred ostension: perception passes through from any one of the tokens of a given type to the type that stands behind it. This view, however, seems to me to be deeply confused: there is no such thing as *the type* that stands behind a token; the definite description that is supposed to denote it uniquely has, in fact, no unique denotation. Let us imagine Quine's letter-token alpha, and let us suppose that it was written with white chalk on a blackboard. The object that we are imagining is not a token of the letter-type alpha only: it is also a token of the white-alpha-type; of the light-coloured-alpha-type; of the alpha-written-with-chalk-type; of the alpha-written-on-a-blackboard-type; of the greek-letter-type; and many,

many more. In Dodd's words, all of these types stand behind Quine's letter-token.[67]

Performances are not special at all in this regard: they, too, instantiate multiple types. Consider, for instance, Ashkenhazy's performance of Beethoven's *Pathétique*. Let S be the set of all possible correct performances of this sonata. Additionally, let S' be the proper subset of S that contains Ashkenazy's 1997 performance, and all of those possible performances whose initial chord is comparable with Ashkenazy's in their loudness. Finally, let S'' be the union of S and the set of all possible sound events that failed to constitute correct performances of the *Pathétique* because they were initiated with a chord too loud or too faint to count as *fortepiano*. Clearly, the *Pathétique* is the type tokened by all and only the members of S. Let K be the type tokened by all and only the members of S'; and K', the type tokened by all and only the members of S''. Then Ashkenazy's performance is a token not only of the *Pathétique*, but of K and K' as well. Performances, then, token a plurality of types; and, if this is the case, the theory that audition passes through from a performance to *the type* that stands behind it is plainly ill-formulated.

If there is anything like *deferred audition*, it must consist of audition passing from a performance to the multiple types that stand behind it; but is there such a thing? Earlier I stated that, on the theory that we are considering, a musical type bears all and only those acoustic properties that are shared by the totality of its correct tokens. The *Pathétique*, then, is characterised by beginning with a *fortepiano* C-minor chord, for all of its correctly formed tokens start with a *fortepiano* C-minor chord. K, whose tokens all bear the property beginning with a frankly sonorous C-minor chord, is characterised by this property as well. Now, all of the tokens of K' begin in C-minor, but the volume of their initial chord varies greatly. Hence, all that can be said about K' is that it is characterised by beginning with a C-minor chord of no particular volume—a C-minor chord *simpliciter*. It follows that the *Pathétique*, K and K' differ in their acoustic qualities. If we hear all of them when we hear Ashkenhazy's 1997 performance of the *Pathétique*, it follows that, when we hear this performance, we hear a multiplicity of things bearing different acoustic qualities—or, to put it more bluntly, a multiplicity of acoustically different things. That is patently *not* what we experience, however, when we hear Ashkenazy's performance: hearing Ashkenazy's performance is nothing like hearing several different pieces of music at once; and, additionally, the idea that

---

67. Cohen (forthcoming) has a neat explanation of how Quine's deferred ostension is possible even when his letter-token alpha instantiates multiple types.

137

we hear multiple things when we hear Ashkenazy's performance becomes even more bewildering when we remind ourselves that the *Pathétique*, K and K' are only three elements of the sheer variety of types that the performance in question instantiates and, thus, three elements of the sheer variety of things that deferred audition should reach through to when hearing Ashkenazy's performance. Moreover, Beethoven composed a musical work that begins with a *fortepiano* C-minor chord; he did not compose a musical work beginning with a frankly sonorous C-minor chord, nor one beginning with a C-minor chord *simpliciter*. If we hear K and K' besides the *Pathétique* when we hear Ashkenazy's performance, then, when we hear this performance, we hear a number of things that Beethoven did not compose—but that, of course, is intuitively not the case at all.

In order to vindicate our intuitions, the advocate of deferred audition could contend, on some grounds or others, that the reason why we do not seem to hear several different things when we hear Ashkenazy's performance is that the *Pathétique*, K and K' all sound, to us, exactly the same as one another, despite differing in their acoustic properties. Surely, though, that cannot be the case. Those who think that musical works are eternally existing universal entities believe that composition is an act of selection (see, e.g., Dodd, 2007, pp. 112–121; Kivy, 1983; 1987). On this view, Beethoven wrote '*fp*' at the very beginning of his authoritative score as part of the process of singling out the *Pathétique* from amongst all the other musical types that exist. In particular, he thereby chose the *Pathétique* in place of K and K'. If the *Pathétique*, K and K' sound the same as one another, however, then that authorial act was entirely futile: there was no reason to choose the *Pathétique* instead of the other types, nor to make a choice at all, because all of the types in question sound exactly the same. I take it, though, that this result is unacceptable: there was definitely a point in Beethoven's decision to write '*fp*'.

Earlier in this section, I submitted that, if musical works are audible, they must be the objects of deferred audition. Every time that we hear a sound, it has a maximally determinate pitch, a maximally determinate volume, and a maximally determinate duration; but musical works have no maximally determinate acoustic properties. I then suggested that, if musical works are audible, it might be because audition passes through from the determinates borne by a performance to the determinables that characterise *the type* that stands behind it. Nevertheless, I argued that this idea is ill-formulated. Perhaps what should be said, instead, is that audition reaches through from a performance to

the various types that stand behind it; but, as I have just shown, this hypothesis delivers untenable results. If my arguments in this section are sound, then not even the ideal musical ontology fosters a construal of musical works as audible. I believe, consequently, that the prospects of success for this construal are rather disheartening.

3   Metonymy

In this section, I will articulate a response to the audibility problem that may be adopted by musical ontologists who believe that the project of construing musical works as audible is daunting, or not very promising. In what follows, I will assume that musical works are abstract and repeatable, as the musical type theorist believes that they are; but I will not commit to any specific view relating to the question of whether musical works possess causal powers or acoustic properties. My proposal, then, may be adopted not only by the musical type theorist, but by whoever identifies musical works with repeatable *abstracta*.

In §1, I argued that the sentences where we seem to say that birds and bells are listened to may be understood as metonymical assertions that certain *sound events* are heard. Similarly, I believe, our sentences about listening to musical works must be understood as metonymical statements concerning performances. This view has it that (17), for instance, is metonymical for (18):

17.   I heard Sibelius's *Finlandia.*
18.   I heard a performance of Sibelius's *Finlandia.*

I think that it is very plausible that we often speak truly about musical works; that we can correctly say, for instance, that they were created by composers, that their form corresponds to a certain category, that they are characteristic of a certain musical tradition, etc. But, on certain occasions, the conclusion seems inevitable that we only *seem* to speak about them; that, sometimes, we speak metonymically about musical works in order to make assertions that concern their performances. Suppose that a, b and c attend a performance of Beethoven's triple concerto and Tchaikovsky's first piano concerto. Whilst they observe the soloists attentively, it becomes apparent that the pianist's technique is quite sloppy, whereas the violinist is remarkably good. Beethoven's triple concerto is brought to a conclusion with a dramatic *crescendo*; no encore is played, and a, b and c leave the auditorium. Then each one of them utters one of the following three sentences:

19.   I have never seen such careless piano playing as that in Tchaikovsky's *cadenza*.

20.   The violinist was the best part of the triple concerto.

21.   The *crescendo* at the very end of the triple concerto was spectacular.

In the hypothetical scenario that I have described, (19)–(21) are all perfectly felicitous; and, intuitively, they are uttered truthfully. Taken literally, however, (19)–(21) concern Beethoven's and Tchaikovsky's musical works; and, if they really concern these works, then they are all utterly false. Clearly, Tchaikovsky's concerto does not feature a passage of careless piano playing; and no particular violinist, nor their playing, is a part of Beethoven's triple concerto. Moreover, the authoritative scores of Beethoven's triple concerto say nothing about a *crescendo* in its final passages, so there is surely no spectacular one at all. If we want to maintain, in accordance with our intuitions, that (19)–(21) were uttered truly, we must regard them as metonymical for (22)–(24):

22.   I have never seen such careless piano playing as that in the performance of Tchaikovsky's *cadenza*.

23.   The violinist was the best part of the performance of the triple concerto.

24.   The *crescendo* at the very end of the performance at triple concerto was spectacular.

It follows that it is uncontroversial that we sometimes talk metonymically about musical works in order to make claims about their performances. Our tendency to economise speech suffices to explain this practice: (19)–(21) are visibly more succinct than (22)–(24). My present contention is merely that our talk about listening to musical works is a further instance of this undoubtedly pervasive phenomenon. My contention, therefore, is substantially supported by our actual linguistic practices.

An obvious objection to this proposal says that it constitutes a distasteful error theory; my response to this objection is that it is simply wrong. The point that I am urging is not that any one of our utterances is untrue: it is merely that some of them—viz., the ones that seem to say that musical works are heard—must not be interpreted at face value. Moreover, the contention that these utterances must not be interpreted at face value is not pre-theoretically counterintuitive at all: upon a little reflection, any layman would happily accept that every time that we engage in the activity of hearing music, we hear either an improvisation or a performance of a musical work—that there are no other pieces of music that we ever listen to. Accordingly, on the face of the question

of whether they have heard anything of Beethoven's ninth symphony that was not a performance thereof, they would respond that they have not.

Solving the audibility problem consists in answering the question of how musical works are appreciated by means of audition if they are abstract. I have argued that musical works are not audible. If this is right, then how are they appreciated by means of audition? Suppose that you attend a piano recital where a work will be performed that you have never heard before. Suppose that you have a strong reason to believe that the work in question may be indeterministic: you have been informed that roughly half of the works of the composer are indeterministic—they consist of very simple themes or motives, and an invitation to performers to develop them into complex pieces. You hear the pianist's performance attentively, and feel very pleased when it is over: you have enjoyed the recital very much. However, you are also very confused: you are still ignorant of whether the work performed is indeterministic or not, and, consequently, you do not know whether the beautiful music you have heard is ultimately an achievement of the composer or the performer. You cannot decide, therefore, whether the work performed is worthy of praise or not.

Most of the time, our interaction with performances of repeatable works is not like this: most of the time, we hear performances not with the conviction that the music that we hear was probably devised, to a significant extent, by the performers themselves—but under the assumption that, generally, performers play the sounds and rhythmic figures that they play because they must do so in order to produce an occurrence of the repeatable object that the relevant musical work is identical to, for these sounds and rhythmic figures *correspond* to the properties that characterise the work performed.[68] I think that we appreciate musical works by means of audition because we possess this simple piece of knowledge. Thanks to it, hearing a performance of a musical work W generally sanctions robust conjectures about the properties that characterize the object that the composer created when they composed W.

On the view that I advocate, then, appreciating a musical work by means of audition involves knowledge about compositional and performative practices. This suggests

---

68. Note that this claim is not committed to the truth or the falsity of the hypothesis that musical works bear acoustic properties. Those who hold this hypothesis may understand the relation of correspondence that I just mentioned as the relation of identity; those who reject the hypothesis in question may understand correspondence as the relation that holds between an acoustic property and the normative property that is associated with it.

that my view goes against the widely held doctrine that aesthetic appreciation is non-inferential; but I do not think that this is too detrimental to my contentions. For one thing, the aforementioned doctrine has been questioned on solid grounds (see, e.g., Dorsch, 2013); and, for another, there are weighty reasons to believe that aesthetic appreciation is inferential in some cases. Suppose that you are hearing a piano performance and that, at the highest point of its climatic passage, the performer plays a chord that you find absolutely astounding. Consequently, you judge the climax of the piece performed to be brilliant. Unbeknownst to you, however, the performer made a mistake: had they played the right chord, you would have found the passage in question to be wholly unremarkable. The reason why you erred in believing that this passage is brilliant has nothing to do with your perceptual abilities: you erred because you mistakenly *believed* that the pianist was playing the sounds that corresponded to the musical work performed. To my mind, the existence of errors of this kind in aesthetic judgement suggests that there are inferential processes involved in some instances of aesthetic appreciation. Finally, in order to do justice to our conviction that we appreciate music directly when we hear a performance, we must simply remember that there is, indeed, a piece of music that is plausibly appreciated non-inferentially when we hear a performance—viz. the performance itself. Some knowledge may be needed in order to judge some of the technical qualities of a performance, but the thinnest of its aesthetic properties may well require no knowledge to be appreciated.

If my arguments in this section are sound, then our discourses relative to the audition of music, as well as the audibility problem, can be both approached in a promising way without supposing that musical works are audible.

4   Conclusion

In§1, and §2, I argued that musical works cannot be construed as audible—not even within an ideal theory of their ontology. In §3, I contended that the sentences where we seem to say that musical works are heard are best understood as metonymical assertions about performances, and, accordingly, that the audibility problem may be satisfactorily solved without attributing audibility to musical works.

In the introduction of this chapter, I argued that it appears plausible that the framework of the musical type theory could be employed to account for the ontology of all the other forms of performative art. I noted, however, that the musical type theory suffers from

two ponderous problems, and one may sensible expect the type-theoretic accounts of the other performing arts to face analogous difficulties. If my arguments in this chapter are sound, then there are satisfactory solutions to both of the problems that haunt the musical type theory: not only to the creatability problem, which has been addressed by Kivy (1983; 1987) and Dodd (2007, pp. 99-142), but to the audibility problem as well. I thus conclude that the framework of the musical type theory may be safely exported to account for performative artworks in general. More specifically: perhaps the type theorist about, say, drama, will encounter a difficulty analogous to the audibility problem: if dramatic works are types, then they are causally inert; they cannot cause perceptual experiences and, in particular, they cannot be seen. This clashes with our intuitions, which have it that we speak truthfully when we say, for instance, that we are planning to see *Hamlet* on the weekend at the local theatre. My metonymical proposal concerning the audibility of musical works can be imported to the dramatic type theory in the obvious manner in order to solve the clash in question.

If my contention above is correct, then types are remarkably promising in the project of assigning, for the sake of parsimony, all abstract artistic creations to the same ontological category. This, I submit, offers further support to the type theory of *ficta* that I defended in Chapter Five. If my arguments above are sound, then the type theory of *ficta* is virtuous not only for the reasons that I presented in Chapter Five, but also because it fits neatly within the promising project of accounting for all abstract artistic creations in type-theoretic terms. If somebody wished to surpass the type theory of *ficta* with an alternative ontology, they would have to demonstrate that their proposal, too, fulfils the above-mentioned *desiderata*—but they would also have to prove that this proposal fosters, as much as its type-theoretic rival, a unified ontology of abstract artistic creations. And, to my mind, such a demonstration would be truly hard to produce. As I said in the introduction to the present chapter, the musical type theory is, nowadays, the most widely held, and no other musical ontology has been developed, defended and perfected as much. If this is the case, then it may well be that the musical type theory is our best available theory of the ontology of musical works. And, if it is, then the project of assigning all abstract artistic creations to one and the same ontological category will demand us to identify *ficta* with types.

This is the end of Part One and, hence, I wish to present a brief conclusion concerning the entirety of the part in question. In §5 in Chapter One, I explained that fictive fictionalists and fictional realists offer competing accounts of fictive statements. Fictive

fictionalists aim at honouring our use of these sentences without postulating *ficta*; fictional realists, by contrast, contend that honouring our use of these sentences *requires* us to posit *ficta*. In Chapter One, I contended that fictive fictionalism is not promising at all. Throughout Chapter Four, Chapter Five and Chapter Six, I have argued that there exists a fictional realist theory that is, by contrast, significantly promising: the type theory. If my arguments are sound, then this theory satisfies the *desiderata* that I formulated in §1 of Chapter Five (cf. §3 in Chapter Five); it is a judicious reaction to the defeat of possibilism (cf. §2 in Chapter Five); it is free from the problems that creationism suffers from (cf. §5 in Chapter Five); and, according to the arguments that I advanced in this chapter, it fits neatly within the promising project of construing a uniform, type-theoretic ontology for all sorts of abstract artistic creations. If all of the above is right, then the type theory of *ficta* is determinately preferable to fictive fictionalism; and fictive fictionalism, consequently, cannot be rationally endorsed.

Two important result follow from this conclusion. First: there is at least one sphere of discourse that realists can handle better than fictionalists—namely, the sphere constituted by fictive statements. And second: as I explained in the Introduction, fictive fictionalism constitutes the theoretical device that fictionalists in general (mathematical, modal, moral fictionalists, etc.) have entrusted to block the failure result. If fictive fictionalism cannot be rationally endorsed, then fictionalists lack a means to protect themselves from said result. In the Introduction, I said that the project that fictionalists in general participate in is the project of encompassing three theoretical virtues: a naturalistic ontology; an anti-revisionary policy concerning our discursive practices; and the ability to foster semantic uniformity across discursive spheres. According to the failure result, the fictionalists' ontology is not naturalistic. Consequently, we may soundly conclude that the fictionalists' project has failed—at least until a device has been developed that effectively blocks the failure result.

According to the type theory of *ficta*, *ficta* are types of concretist non-actual individuals. They are, therefore, abstract, and their identity conditions ought to be formulated in terms of non-actual individuals. On the type theory of *ficta*, then, *ficta* are doubly exotic. Now, according to the arguments that I have thus far advanced, it is the type theory of *ficta*, and not fictive fictionalism, that comprises a satisfactory account of fictive statements. If my arguments are sound, then we are, by our mere use of fictive statements, committed to the existence not of one, but of two sorts of exotic entities. As I explained in the Introduction, two interesting results follow from this conclusion.

144

The first is that mathematical fictionalism and modal fictionalism are futile. As I will explain in Chapter Seven and Chapter Eight, mathematical and modal fictionalists wish to account for mathematical and modal discourses in fictionalist terms in order to deliver us from committing, respectively, to *abstracta* and concretist non-actual individuals. If *ficta* are types of concretist non-actual individuals, however, then we are inescapably committed to both concretist non-actual individuals and *abstracta* merely in virtue of our use of fictive statements. The second result concerns not mathematical and modal fictionalists only, but fictionalists in general. If in virtue of our use of fictive statements we are committed to two sorts of entities that are non-naturalistic, then, to my mind, the question arises of whether it is worth even trying to escape other sorts of non-naturalistic commitments. Is it worth, for instance, trying to escape a commitment to moral properties on the grounds that they are non-naturalistic if we will have non-naturalistic commitments anyway? Or might it be best to call into question our naturalistic convictions?

I shall not attempt to answer these questions in my thesis. Instead, I will leave them open as potentially interesting paths of inquiry. In the following part of my thesis, I will discuss three distinct fictionalist theories—mathematical fictionalism, modal fictionalism, and moral fictionalism—and I will advance a novel objection against each one of them. Importantly, moreover, in §5 of Chapter Nine, I will show that the core contention behind each one of these three objections is one and the same.

Part Two

Chapter Seven: Higher-Order Fictionalism and the Infinite Regress

The term 'fictionalism' has been used in the philosophy of mathematics to refer to a multitude of theories that differ greatly from one another. A prominent distinction that exists amongst them tells apart Fieldian and post-Fieldian theories (also called 'hard-road' and 'easy-road' theories, respectively). Fieldian theories (see, e.g., Field, 1980), however, have proved to be objectionable on multiple grounds (see, e.g. Malament, 1982; Shapiro, 1983; Resnik, 1985; Chihara, 1990, pp. 46–81), so many fictionalists nowadays believe that the prospects for success of the fictionalist project depend on the post-Fieldian alternatives (see, e.g., Balaguer, 1998, pp. 93–179; Melia, 2000; Yablo, 2001; 2005; 2012; Leng, 2010). This chapter is about them and, consequently, in what follows, I will use the term 'fictionalist theories' to refer to post-Fieldian proposals exclusively.

Consider:

1.   A hobbit destroyed the Ring of Power.

(1) involves existential quantification over hobbits, and purported reference to the Ring of Power. (1), then, is untrue: hobbits and the Ring of Power do not exist. Ordinarily, however, (1) is uttered in order to assert not that which is literally expressed by (1), but the following:

2.   According to *The Lord of the Rings*, a hobbit destroyed the Ring of Power.

Asserting (2) does not commit us to the existence of hobbits or the Ring of Power: it commits us only to the existence of a work of fiction called '*The Lord of the Rings*',

according to which a hobbit destroyed the Ring of Power. If ordinary utterances of (1) express (2), then these utterances do not involve an ontological commitment to hobbits nor the Ring of Power. (2), moreover, is true. If ordinary utterances of (1) express (2), then these utterances are truthful.

Mathematical fictionalism (which I will call simply 'fictionalism' in this chapter) is the view that mathematical statements are *useful fictions*—very much, that is to say, like (1): they are untrue, and yet their utterances are truthful and free of problematic ontological commitments. Witness:

    3.   2 is the only number that is both even and prime.

Fictionalists believe that (3) involves existential quantification over mathematical objects, and purported reference to a mathematical object. They also believe, however, that mathematical objects do not exist. On their view, then, (3) is untrue. Nevertheless, fictionalists contend also that (3) is ordinarily used to assert not that which (3) literally expresses, but the following:

    4.   According to pure mathematical theories, 2 is the only number that is both even and prime.

Asserting (4) does not commit us to the existence of any mathematical objects: it commits us only to the existence of pure mathematical theories that state that 2 is the only number that is both even and prime. If ordinary utterances of (3) express (4), then these utterances do not involve an ontological commitment to mathematical objects.[69] (4), moreover, is true. If ordinary utterances of (3) express (4), then the utterances in question are truthful.

In general and more perspicuous terms, then, fictionalist theories may be said to be characterised by encompassing the following three doctrines:

---

69. One might argue that asserting (4) does commit us to mathematical objects: it commits us to the existence of theories, and theories are mathematical objects: on their linguistic conception, they are sets of propositions or sentence-types; on their semantic conception, they are sets of models. In either case, theories are sets, and sets are mathematical objects, so asserting (4) does commit us to mathematical objects. This chapter is about an objection that has been raised against fictionalist theories and goes precisely along the lines of the aforestated argument: it states that fictionalists are inadvertently committed to the entities that they were set to reject.

Anti-Realism

Mathematical objects do not exist.

Standard Semantics

Mathematical sentences ought to be interpreted at face value: as their surface form suggests, they involve purported reference to mathematical objects, or existential quantification over mathematical objects. Since mathematical objects do not exist, mathematical sentences are untrue.[70]

Anti-Revisionism

For every mathematical statement s, s is typically not used to assert that which s literally expresses. Instead, s is ordinarily used non-literally to convey a piece of veridical information about the non-mathematical realm. Ordinary utterances of mathematical statements, then, are truthful, and free of ontological commitments to mathematical objects. Consequently, we may rationally use mathematical statements as we always have—even if they are untrue.

Fictionalist theories seem to be superior to all of their rivals in so far as they encompass the aforestated doctrines. (Anti-Realism) appears superior to Platonism because it is ontologically less expensive. Some theories attribute non-standard semantic properties to mathematical statements in order to dissociate them from undesirable ontological commitments, and these theories are all inferior to (Standard Semantics) insofar as (Standard Semantics), but not the theories in question, fosters semantic uniformity across mathematics and all other spheres of discourse. Finally, (Anti-Revisionism) appears superior to error theories because it allows us to rationally preserve our current discursive practices whilst upholding an anti-realist stance towards mathematical objects. The fictionalist project, in sum, consists in articulating a theory about mathematics that possesses three notable theoretical virtues: ontological parsimony, anti-revisionism, and aptness for fostering semantic uniformity.

The fictionalist project is, therefore, commendable; but fictionalist theories are all subject to a charge that has, I believe, not received enough attention yet: the nominalistic

---

70. Importantly, the fictionalist postulate that mathematical sentences are untrue is not a universally quantified statement. According to pure mathematical theories, there does not exist a largest prime number. Fictionalists agree: on their view, there does not exist a prime number that is the largest because numbers do not exist at all. It follows that some of the sentences of pure mathematics are true within the fictionalist framework.

impurity objection. Fictionalists generally reject the existence of mathematical objects on nominalistic grounds: they reject them, in other words, because they reject the realm of the abstract. Like other nominalists, fictionalists adduce naturalistic intuitions, epistemological plights and considerations of parsimony in order to deny that abstract entities exist. If the nominalistic impurity objection is sound, however, then fictionalism and nominalism are inconsistent. (Standard Semantics) states that mathematical sentences are untrue. Accordingly, fictionalists claim that (3), for instance, is untrue. Clearly, the subject matter of this claim is not a sentence-token: that which fictionalists claim to be untrue is not any particular utterance, nor any particular string of ink marks. Hence, the subject matter of the claim in question is a sentence-type. Fictionalists themselves explicitly accept a generalisation of this conclusion: fictionalist theories, they declare, concern not sentence-tokens, but sentence-types (see, e.g., Balaguer, 1998, p. 13). Types are abstract, however, so fictionalists inadvertently affirm the existence of precisely those entities that their advocates are set to reject. Adhering to a fictionalist theory, in sum, is futile: it amounts to exchanging an ontological commitment with certain sorts of *abstracta* for an ontological commitment with other sorts thereof.[71]

Balaguer (1998, p. 13) acknowledges this futility result, and recommends fictionalists to block it by endorsing a second-order fictionalist theory—a fictionalist theory, that is to say, about fictionalist theories. According to second-order fictionalism, fictionalist theories are useful fictions: they are untrue because they affirm the existence of sentence-types, and sentence types do not exist—but they are informative nonetheless.

Prima facie, Balaguer's strategy to resist the nominalistic impurity objection appears theoretically commendable: it allows fictionalists to expel sentence-types from their ontology by way of the same semantically conservative, non-revisionary method that

---

71. There is an interesting reply to the nominalistic impurity objection that suggests itself immediately: sentence-types and mathematical objects are all abstract, but they belong to two different classes of abstract entities. Fictionalists, the replier claims, reject the existence of the members of the class that mathematical objects pertain to, but accept the existence of the class where sentence-types belong. Thus, the replier concludes, fictionalists are not, after all, committed to the existence of the entities that they are set to reject. In footnote 69, however, I showed that it is easy to contend, with significant plausibility, that fictionalists are inadvertently committed to sets. The nominalistic impurity objection, then, could be reformulated so as to appeal to the fictionalists' inadvertent commitment not to sentence-types, but to incontrovertibly mathematical objects. On this reformulation, the conclusion is inescapable that fictionalists are committed to *exactly* those entities that they wished to reject. For considerations of simplicity that may not be apparent, though, I will discuss, in what follows, only the sentence-type version of the nominalistic impurity objection.

they used to exclude other sorts of *abstracta* before. Nevertheless, Daly (2008, pp. 433–440) maintains that the strategy in question is ineffective: the postulation of a second-order fictionalist theory, he contends, produces an infinite regress that precludes fictionalists from denying the existence of abstract entities. In this chapter, I will contest Daly's contention: I will argue that the postulation of a second-order fictionalist theory does indeed generate an infinite regress, but one that is innocuous for the nominalistic purity of fictionalist theories. Afterwards, however, I will argue that the existence of the infinite regress signifies that fictionalists cannot uphold an anti-revisionist stance in relation to mathematical statements. Finally, I will contend that the loss of the anti-revisionism postulate is catastrophic for the fictionalist project.

The purpose of this chapter is twofold. I wish to show that the reasons why Balaguer's strategy is unsuccessful are not those that Daly thought, and I wish to do so not in order to counter Daly, but in order to establish soundly that the strategy in question is, indeed, unsuccessful. Moreover, I wish to establish that Balaguer's strategy fails in order to draw renewed attention to the nominalistic impurity objection: I believe that this objection poses a significant threat to the fictionalist project and, accordingly, that it deserves more serious consideration than it has thus far received.

1   Two Rejoinders to the Nominalistic Impurity Objection

Let us analyse a logically tight reconstruction of the nominalistic impurity objection in order to investigate how it could be countered. Let us first note that (3) is a mathematical sentence. Now consider:

> 5.   (3) is an untrue mathematical sentence.
> 6.   At least one mathematical sentence is untrue.

(5) and (6) are postulates of any fictionalist theory. (5) follows from (Standard Semantics) and (Anti-Realism): if mathematical sentences are to be read at face value, then (3) quantifies existentially over mathematical objects, and purports to refer to a mathematical object. Moreover, (3) is untrue because there are no mathematical objects. It follows that (5) is true. (6), in turn, follows from (5). In sum: if fictionalists endorse (Standard Semantics) and (Anti-Realism), they ought to endorse (5) and (6). (5) and (6) appear to have, respectively, the following logical forms: 'Φa' and '$\exists x \Phi x$'. If these are indeed the logical forms of (5) and (6), then (5) involves purported reference to a sentence-type, and (6) quantifies existentially over sentence-types. Sentence-types are

abstract. Fictionalist theories, then, involve purported reference to abstract objects, and existential quantification over abstract objects. Since fictionalists are committed to the truth of fictionalist theories, they are committed to the existence of *abstracta*.

Thus goes the nominalistic impurity objection. How could fictionalists reject its soundness? It seems to me that there are only two steps within my reconstruction of the nominalistic impurity objection that could be plausibly disputed. The first is the step where the surface forms of (5) and (6) are supposed to be sufficient to infer validly that (5) and (6) involve, respectively, purported reference to sentence-types, and existential quantification over sentence-types. The second is the step where fictionalists are accused of committing to the existence of *abstracta* on the grounds that they are committed to the truth of fictionalist theories. If this is right, then there are two doctrines that fictionalists could put forward in order to block the nominalistic impurity objection:

Paraphrase Type Nominalism
Fictionalist postulates appear to involve purported reference to sentence-types, and existential quantification over sentence-types, but, in reality, they do not.

Deflationary-Truth Type Nominalism
Fictionalist postulates involve purported reference to sentence-types, and existential quantification over sentence-types, and sentence-types do not exist, but fictionalist postulates are true nonetheless.

Suppose that fictionalists endorse (Paraphrase Type Nominalism). Then they deny that fictionalist postulates like (5) and (6) involve purported reference to sentence-types, or existential quantification over sentence-types. Perhaps, on their view, the surface form of these postulates is deceiving; perhaps purported reference to sentence-types and existential quantification over sentence-types disappear on analysis. Perhaps these postulates disguisedly concern sentence-tokens only. If this is the case, then committing to the truth of a fictionalist theory does not commit fictionalists to the existence of sentence-types.

Suppose, alternatively, that fictionalists endorse (Deflationary-Truth Type Nominalism). They, then, believe that fictionalist postulates purport to refer to sentence-types, and quantify existentially over sentence-types, and they also believe that sentence-types do not exist. On their view, however, fictionalist postulates are true nonetheless. Con-

sequently, committing to the truth of a fictionalist theory does not commit fictionalists to the existence of sentence-types.

Adhering to either of the doctrines in question, in sum, would allow fictionalists to avow the truth of their theories without committing to the existence of abstract objects. Adhering to either of these doctrines, though, comes at a non-negligible cost—the cost, that is to say, of taking on a non-standard account of the semantics of fictionalist postulates: (Paraphrase Type Nominalism) constitutes a non-standard account of the logical form of fictionalist postulates; (Deflationary-Truth Type Nominalism) constitutes a non-standard account of their truth conditions.

Taking on a non-standard account of the semantics of a certain sphere of discourse is not problematic *per se:* if one is willing to renounce semantic uniformity across the various extant discursive spheres, then there is not much that can be said against one's deviant semantic convictions. I submit, however, that non-standard semantic doctrines are problematic within the framework of the fictionalist project. Fictionalists, we have seen, endorse (Standard Semantics); they endorse, that is to say, a standard account of the semantics of mathematical statements. On this account, mathematical statements possess standard logical forms and standard truth-conditions. As I said earlier, (Standard Semantics) allows fictionalists to claim superiority over those of their rivals who attribute deviant semantic properties to mathematical statements. Fictionalists seem superior to these rivals because, unlike the rivals in question, fictionalists appear to uphold semantic uniformity across all discursive spheres. Obviously, however, they do not uphold universal semantic uniformity at all if the theory that they adhere to is itself a semantically non-standard sphere of discourse. If fictionalism is semantically non-standard, that is to say, fictionalists may attribute a standard semantics to the discursive sphere of mathematics only at the cost of generating a new sphere of discourse (viz., fictionalism) that is not semantically standard and, therefore, at the cost of shattering semantic uniformity across discursive spheres. Endorsing either (Paraphrase Type Nominalism) or (Deflationary-Truth Type Nominalism), then, forces fictionalists to renounce the theoretical virtue that positively distinguished them from a number of alternative doctrines. If fictionalism does not possess this virtue after all, then fictionalists ought to tell us why their theory is preferable to the relevant alternatives. More importantly, however, if fictionalism is itself semantically deviant, then the fictionalist project simply cannot be carried through—for this project, recall, consisted in articulating an account of mathematics that fosters not only ontological parsimony and

anti-revisionism, but semantic uniformity as well.

If my arguments above are sound, then neither (Paraphrase Type Nominalism) nor (Deflationary-Truth Type Nominalism) are available to fictionalists. If they wish to counter the nominalistic impurity objection, therefore, they must follow Balaguer and posit a second-order fictionalist theory.

2   Daly's Dilemma

The nominalistic impurity objection states that fictionalist theories are nominalistically impure: they affirm the existence of abstract sentence-types and, thus, they commit their advocates to precisely those entities that they wished to reject. Balaguer, recall, proposes to counter this objection by positing a second-order fictionalist theory. This theory states that fictionalist postulates are useful fictions: they are untrue, but ordinary utterances of them are informative nonetheless. In more perspicuous terms: second-order fictionalism encompasses the following three doctrines:

Anti-Realism$_2$
Sentence-types do not exist.

Standard Semantics$_2$
Fictionalist postulates ought to be interpreted at face value: as their surface form suggests, they involve purported reference to sentence-types, or existential quantification over sentence-types. Since sentence-types do not exist, fictionalist postulates are untrue.

Anti-Revisionism$_2$
For every fictionalist postulate s, s is typically not used to assert that which s literally expresses. Instead, s is ordinarily used non-literally to convey a piece of veridical information about the non-abstract realm. Ordinary utterances of fictionalist postulates, then, are truthful, and free of ontological commitments to sentence-types. Consequently, we may rationally utter and assent to fictionalist postulates—even if they are all untrue.

(Standard Semantics$_2$) states that the surface form of fictionalist postulates reveals their logical form: fictionalist postulates do not only appear to purport to refer to sentence-types and quantify over sentence-types, but do indeed do so. (Standard

Semantics$_2$) also states that fictionalist postulates are untrue because sentence-types do not exist. Hence, second-order fictionalism, unlike (Paraphrase Type Nominalism) and (Deflationary-Truth Type Nominalism), comprises a standard account of both the logical form and the truth conditions of fictionalist postulates. Second-order fictionalism, that is to say, construes fictionalist postulates as semantically standard and, consequently, as consistent with the fictionalists' goal of establishing semantic uniformity across all discursive spheres. Second-order fictionalism, in sum, possess precisely that theoretical virtue which (Paraphrase Type Nominalism) and (Deflationary-Truth Type Nominalism) lack.

Unfortunately, second-order fictionalism also possesses the problematic quality that characterises its first-order counterparts: it commits its advocates to the existence of abstract objects. (5), recall, is a fictionalist postulate. It appears to involve purported reference to sentence-types; by (Standard Semantics$_2$), it does involve purported reference to sentence-types. According to (Anti-Realism$_2$), sentence-types do not exist. (Standard Semantics$_2$) thus entails that (5) is untrue. The following, then, is a postulate of second-order fictionalism:

7.   (5) is untrue.

Clearly, the term '(5)' in (7) does not refer to a sentence-token: it is not a particular utterance nor particular string of ink marks that is untrue according to (7). Consequently, '(5)' refers to a sentence-type. Those who endorse second-order fictionalism, therefore, are committed to the truth of (7), and those who are committed to the truth of (7) are committed to the existence of abstract objects.

In order to avoid this result, fictionalists ought to postulate a third-order fictionalist theory—a fictionalist theory, that is to say, about second-order fictionalism. This, Daly (2008, pp. 435–436) contends, implies that the postulation of a second-order fictionalist theory yields an infinite regress. Third-order fictionalism states that the postulates of second-order fictionalism are useful fictions. Like fictionalist postulates, the postulates of second-order fictionalism are not sentence-tokens: they are not particular utterances, nor particular strings of ink marks. Third-order fictionalism, then, affirms the existence of sentence-types, and, consequently, it commits its advocates to the existence of abstract objects. In order to elude this unwanted commitment, fictionalists must posit a fictionalist theory of an even higher order—and this process continues *ad infinitum*. In sum: if fictionalists are to remain nominalistic, they are bound to accept that, for every

fictionalist theory T, there is a fictionalist theory about T that allows them to avoid the unacceptable commitments otherwise brought by T.

Daly (2008, pp. 433–437) argues that the existence of the infinite hierarchy of fictionalist theories forces fictionalists to commit to abstract objects. His argument is a dilemma: the infinite hierarchy of fictionalist theories is either actual or potential. If it is actual, then fictionalists are committed to the existence of infinitely many theory-types: finite humans will never be able to produce an infinite number of theory-tokens, so the infinitely many theories that constitute the interminable hierarchy must be theory-types. If the hierarchy is potential instead, then fictionalists are committed to the existence of merely possible theories. In either case, Daly concludes, fictionalists are committed to precisely those entities that they wished to reject.[72]

I agree with Daly that the postulation of a second-order fictionalist theory yields an infinite regress, and I am very sympathetic to his contention that this regress is detrimental to the fictionalist project. I think, however, that his dilemma is unsound. Consider a fictionalist who adhered to Balaguer's strategy to counter the nominalistic impurity objection: she endorsed a second-order fictionalist theory. Suppose that she eventually realised that second-order fictionalism produces an infinite regress. She now admits that, for every fictionalist theory T, there is a theory about T that states that T is a useful fiction. If this is the case, then she ought to accept that fictive discourse is extremely pervasive: on her view, the best theory that there is about mathematical sentences states that all of these sentences are fictive; this theory, moreover, is best accounted for by a second-order theory that states that the first-order theory in question is fictive, too, and this second-order theory is best understood as a piece of fictive language as well—and so on. Once the fictionalist has reached this point, I believe, there is no reason why she should take seriously the statement that there is an infinite hierarchy of fictionalist theories. I think, in other words, that she may well reply to

---

72. Clearly, Daly believes that merely possible entities constitute a subset of the set of abstract entities: if, by committing to merely possible theories, fictionalists commit to the entities whose existence they want to deny, then merely possible theories must be abstract. Lewis (1986), though, famously contends that merely possible entities are concrete. Daly (2008, p. 436) discards this view by claiming that merely possible entities are abstract because they do not exist in actual spacetime. Thus, on Daly's view, abstract entities are those that do not exist in actual spacetime, rather than those that do not exist in any region of spacetime at all—neither actual nor non-actual. Perhaps, however, the entities that the nominalist repudiates are not those that do not exist in actual spacetime, but those that do not exist in any region of spacetime at all. If this is the case, then Daly's dilemma will not lead the nominalist to commit to the entities that she wants to reject: it will lead her to commit to entities that exist outside *our* spacetime, but not to entities that are non-spatiotemporal *tout court*.

Daly that the statement in question is nothing more than a useful fiction that carries no undesirable ontological commitments.

Daly argues that the theories that constitute the infinite hierarchy ought to be abstract. He thereby concludes that fictionalists are committed to the entities that they were set to reject. But there is no reason why fictionalists should not avoid committing to abstract theories by way of the very procedure that they have used to avoid committing to other sorts of *abstracta* before. They avoided committing to mathematical objects by claiming that the statements that affirm the existence of mathematical objects are useful fictions; they avoided committing to sentence-types by claiming that the sentences that affirm the existence of sentence-types are useful fictions; and they may, analogously, avoid committing to abstract theories by claiming that the sentences that affirm the existence of abstract theories are useful fictions. Clearly, moreover, the fact that the theories that constitute the fictionalist's hierarchy are infinitely many cannot preclude the fictionalist from refusing to commit to them: she has already succeeded in eluding an ontological commitment to infinitely many numbers and infinitely many sets.

If my arguments above are sound, then, *pace* Daly, the infinite regress of fictionalist theories does not lead fictionalists to an ontological commitment to abstract objects. We may then wish to conclude that Balaguer's response to the nominalistic impurity objection is successful; that it does generate an infinite regress, but one that is innocuous for the fictionalists' commitments. In the following section, however, I will contend that this conclusion is unsound: the infinite regress, I will argue, does not lead fictionalists to an ontological commitment to abstract objects, but it leads them nonetheless to the collapse of the fictionalist project.

3   The Infinite Regress

Fictionalists, we have seen, all agree that, for every mathematical statement s, s is typically not used to assert that which s literally expresses. Instead, s is ordinarily used non-literally to convey a piece of veridical information about the non-mathematical realm. Fictionalists, though, hold diverging views concerning the mechanisms by means of which the ordinary utterances of mathematical statements convey truthful information. A notable account of the mechanisms in question has been drawn from Walton's work on pretence and the representational arts (1990). According to Walton, every game of make-believe is governed by *principles of generation*—principles, that is to say,

that determine what is the case within the relevant game. The fencing game, for instance, is governed by a principle that states that all the players are, within the game, skilful fencers. Some principles of generation, moreover, are *bridging principles:* principles that state that a certain sentence s_1 counts as true within the game just in case another sentence s_2 is true in reality. The fencing game, for example, is governed by the following bridging principle:

Fencing

'x caused a deadly injury to y' is true within the game just in case 'x touched y' is true in the real world.

Walton thinks that we have an intuitive understanding of the bridging principles that govern the make-believe games that we participate in. As a consequence, assertions made within these games convey truthful information about the real world at the level of pragmatics. Suppose, for example, that a is participating in the fencing game, and touches b. She triumphantly exclaims that she has caused a deadly injury to b. We all understand that the fencing game is governed by (Fencing) and, consequently, a's exclamation pragmatically conveys that 'a touched b' is true in the real world.

Yablo (2005, pp. 98–99) and Leng (2010, pp. 169–180) suggest that ordinary utterances of mathematical statements are akin to the utterances that are occur within make-believe games. Yablo maintains that, for every pure mathematical statement s, the ordinary utterances of s convey, by the Waltonian mechanism outlined above, that the following sentence is true in the real world: 'According to the theories of pure mathematics, s'. On this view, ordinary utterances of (3), for instance, express that (4) is true in the real world. Moreover, Yablo and Leng both contend that ordinary utterances of applied mathematical statements convey, by Walton's mechanism, various pieces of veridical information about the non-mathematical realm. Consider, for example:

8.   The number of Martian moons is 2.

On the fictionalist's view, (8) involves existential quantification over mathematical objects, and purported reference to mathematical objects, and it is, therefore, untrue. Yablo and Leng, though, believe that (8) is typically not used to assert that which is literally expressed by it. Instead, they claim, (8) is used non-literally to affirm that the following sentence is true in the real world:

9. $\exists x \exists y (x$ is a martian moon $\land y$ is a martian moon $\land x \neq y \land \forall z (z$ is a martian moon $\rightarrow z = x \lor z = y))$

Several fictionalists disagree with both Yablo and Leng, however. Field (1989, pp. 1–8) and Balaguer (1998, p. 13), for instance, believe, much like Leng and Yablo, that, for every pure mathematical statement s, s is typically used to express that, according to pure mathematical theories, s. Nevertheless, Balaguer and Field maintain that the mechanism whereby s does this is not the mechanism hypothesised by Walton, but simple, ordinary ellipsis: s, that is to say, is typically asserted as governed by an elided sentential operator. Thus Balaguer and Field believe that (3), for instance, is typically uttered as elliptical for (4).

In an earlier work, moreover, Yablo (2001) compares ordinary utterances of applied mathematical statements not with the utterances that occur within make-believe games, but with metaphors. A metaphorical statement, Yablo suggests, has both a *literal content* and a *real content:* the former is wholly determined by the meaning of the expressions that constitute the statement in question; the latter is a piece of information that this statement is ordinarily used to non-literally convey. Consider, for instance:

10. Juliet is the sun.

The literal content of (10) is the proposition that Juliet and the sun are numerically identical; its real content, by contrast, is that Juliet bears some of the properties that characterise the sun— like the property of being dazzling, perhaps. Yablo believes that applied mathematical statements, like metaphorical statements, possess a literal content and a real content, and they are ordinarily used to assert not the former, but the latter. On his view, (8), for example, has a literal content and a real content. The former is the proposition that there exists exactly one number associated to the Martian moons, and that number is numerically identical to 2; the latter is represented by (9). Moreover, (8) is ordinarily used to assert not its literal content, but its real content.

Similarly, Balaguer (1998, pp. 132–136), Melia (2000, pp. 466–475) and a later time slice of Yablo (2012, pp. 1009–1011) suggest in varied ways that a mathematical statement has both a literal content and a *purely nominalistic content.* Its literal content is, again, wholly determined by the meaning of the expressions that constitute it; its purely nominalistic content, by contrast, is a proposition that is non-literally expressed by its ordinary utterances, and concerns the physical world exclusively—and not, that is to

say, the realm of the abstract at all. The literal content of (8), for instance, is, again, the proposition that there exists exactly one number associated to the Martian moons, and that number is numerically identical to 2; its nominalistic content is represented by (9). Ordinarily, these three authors maintain, when we put forward a mathematical statement, what we affirm and commit to is not its literal content, but its nominalistic content only.

I shall not discuss these various proposals in any further detail: my present aim consists only in demonstrating that fictionalists agree with one another that ordinary utterances of mathematical statements convey truthful information—even if they disagree with regard to the mechanisms by means of which this information is conveyed. Fictionalists, then, adhere all of them to the following principle:

Informativeness
For every mathematical statement s, s is typically not used to assert that which s literally expresses. Instead, s is ordinarily used non-literally in order to convey a piece of veridical information about the non-mathematical realm.

(Informativeness) allows fictionalists to contend that ordinary utterances of mathematical statements are truthful, and do not involve an ontological commitment to mathematical objects. That the utterances of mathematical statements are truthful and free of pernicious ontological commitments, in turn, allows fictionalists to advance their anti-revisionist doctrine: if mathematical statements are ordinarily used informatively, we may rationally put them forward as we always have—even if they are literally untrue. It is (Informativeness), in short, that allows fictionalists to sustain their anti-revisionist convictions. In what follows, however, I will argue that the fictionalists' adherence to an infinite hierarchy of fictionalist theories precludes them from upholding (Informativeness).

According to second-order fictionalism, the postulates of first-order fictionalist theories are untrue, but informative. Since (Informativeness) is a postulate of first-order fictionalist theories, (Informativeness) is untrue and informative. According to third-order fictionalism, however, second-order fictionalism is not true—so it is not true that (Informativeness) is informative. Third-order fictionalism states that second-order fictionalism is informative, though—so, according to third-order fictionalism, the second-order postulate that (Informativeness) is informative is informative. Nevertheless, according to fourth-order fictionalism, third-order fictionalism is not true. Thus, it is not true,

after all, that the postulate that (Informativeness) is informative is informative. This process continues endlessly. No matter how far we ascend along the infinite hierarchy of fictionalist theories, we will never find a true statement saying that there is information expressed by the statement that says that there is information that is expressed by that statement that says... that there is information expressed by (Informativeness). At every step along the infinite regress of fictionalist theories, we will discover that what we had learned at our previous step about there being some trace of informativeness in (Informativeness) is, in fact, untrue. If this is the case, however, then there is no trace of informativeness in (Informativeness) at all: it is not informative; the sentence that says that it is informative is not informative; there is a statement that says that the sentence that states that (Informativeness) is informative is informative, and that statement is not informative; etcetera.

(Informativeness) is a postulate of first-order fictionalist theories. Second-order fictionalism states that fist-order fictionalist theories are untrue. (Informativeness), then, is untrue. Moreover, if my arguments above are sound, (Informativeness) is also thoroughly uninformative: it is devoid of any trace of informativeness. There is, then, nothing veridical that we express, literally or non-literally, when we assert that mathematical sentences are ordinarily used to convey truthful information. Mathematical sentences, simply put, are plainly uninformative.

Fictionalists maintain that mathematical sentences are untrue. If they are, as I have argued, also plainly uninformative, it is irrational to put them forward as we ordinarily do so: there is nothing truthful that we express, literally or non-literally, when we assert them. Thus the fictionalists' anti-revisionism doctrine collapses: rationality demands us not to use mathematical sentences as we usually do.

In §1, I explained that the fictionalist project consists in articulating a theory about mathematics that encompasses semantic uniformity, ontological parsimony, and anti-revisionism. In this section, I have argued that positing an infinite hierarchy of fictionalist theories precludes fictionalists from upholding an anti-revisionary stance towards mathematical sentences. If my arguments are sound, then positing this hierarchy radically prevents fictionalists from completing the fictionalist project.

In §2, I argued that Balaguer's strategy to counter the nominalistic impurity objection is superior to its most obvious alternatives. Fictionalists, I concluded, had better posit a second-order fictionalist theory. In §3, I reconstructed Daly's argument to the effect that

161

the postulation of a second-order fictionalist theory produces an infinite regress and, thus, forces fictionalists to posit an infinite hierarchy of fictionalist theories. In this section, I contended that positing an infinite hierarchy of fictionalists theories precludes fictionalists from endorsing (Anti-Revisionism). If the arguments that I have presented thus far are sound, then all that fictionalists can do is to adhere to (Anti-Realism) and (Standard Semantics). They can claim, that is to say, that mathematical objects do not exist, and they can claim that mathematical sentences are untrue because they involve purported reference to mathematical objects, and existential quantification over mathematical objects—but that is all that they can claim. Nevertheless, a doctrine that is exhaustively characterised by (Anti-Realism) and (Standard Semantics) is nothing more than a crude error theory: it is a theory that states that one of our discursive spheres is flawed because it is filled with plainly untrue statements. And, in order to advance an error theory, one need not to appeal to any of the apparatus that constitute the fictionalists' theoretical machinery: one need not appeal to any devices for non-literal information transmission, nor any make-believe games, Waltonian mechanisms, ellipses, metaphors, literal contents, real contents, or purely nominalistic contents, nor need one posit an infinitely large structure of hierarchically ordered theories. Thus it turns out that the whole, interminable hierarchy of fictionalist theories, as well as the the entirety of the fictionalists' theoretical machinery, is absolutely futile: it cannot help fictionalists to surpass, even slightly, the crudest anti-realist.

4   Conclusion

By way of conclusion, I wish to contend that the nominalistic impurity objection deserves more serious consideration than it has thus far received: if the arguments that I have presented in this chapter are sound, then it constitutes a major threat to the fictionalist project. In the introduction of this chapter, I said that the fictionalist project consists in articulating a theory about mathematics that encompasses three commendable doctrines: ontological conservativeness, semantic uniformity, and anti-revisionism. The nominalistic impurity objection threatens the first of these doctrines: it states that fictionalists are inadvertently committed to abstract entities. In §1, I assessed two potential rejoinders to the nominalistic impurity objection: (Paraphrase Type Nominalism) and (Deflationary-Truth Type Nominalism). Then I showed that endorsing either one of these rejoinders leads fictionalists to betray semantic uniformity. Then, in §2 and §3, I argued that Balaguer's strategy to counter the nominalistic impurity objec-

tion produces an infinite regress that, ultimately, precludes fictionalists from sustaining their anti-revisionist convictions. The fictionalist project, I thus conclude, is doomed: the three doctrines that fictionalists wish to combine cannot be all of them upheld at once on the face of the nominalistic impurity objection.

Chapter Eight:
The Fiction of Possible Worlds

Modal fictionalists believe that modal statements express propositions concerning the contents of *the fiction of possible worlds*—but what fiction is this fiction? Amongst the wealth of stories that could be composed about alternative universes, which one is the fictionalist's fiction? Most modal fictionalists have never addressed this question (Nolan, 2022, §1.1), but a number of competing answers have been advanced nonetheless by Armstrong (1989, pp. 37–118), Rosen (1990, pp. 333–335, 349–354), Woodward (2011, pp. 537–541) and Nolan (2022, §4.3). In this chapter, I will argue that the *strong* modal fictionalist,[73] who wishes to provide a reductive analysis of modal discourse, cannot accept any one of these answers: they are all ultimately non-reductive, self-defeating, or subject to the well-known fetishism charge (see, e.g., Rosen, 1990, pp. 349–354; Peacocke, 1999, pp. 144–155; Woodward, 2011, pp. 535–541). If my arguments are sound, strong modal fictionalists still owe us an answer: what fiction is their fiction of possible worlds? Unless an adequate response is given, it will remain unclear what, on their theory, modal statements are about.

In §1, I will motivate strong modal fictionalism and, in §2, I will introduce its basic tenets. In §3, I will assess Rosen's, Woodward's and Nolan's responses to the question of what fiction is the fictionalist's fiction of possible worlds. §4 will be devoted to Armstrong's response.

1   Possible Worlds

The following is the possible worlds analysis of modal discourse:

---

73. The distinction between strong and timid modal fictionalists was originally drawn by Rosen (1990, pp. 354). The former are those who strive for a reductive analysis of modal discourse; the latter are those who do not.

Possible Worlds

Possibly $p \leftrightarrow$ there is a possible world w such that, at w, p.

Necessity is defined as the dual by negation of possibility. In the possible world analysis of modal discourse, then, modal operators are interpreted as quantifiers ranging over possible worlds—but what are possible worlds? Many answers have been offered to this question and I do not intend to review them all here.[74] I wish only to point out that some of them foster a *reductive* analysis of modal discourse, whilst others do not.

Abstractionists (e.g., Adams, 1974) adhere to (Possible Worlds), and contend that a possible world is a maximally consistent set of propositions. Consistency, though, is standardly defined as a modal notion, so abstractionists define modal notions in terms of possible worlds, and possible worlds, in turn, in terms of a modal notion. Concretists (e.g., Lewis, 1986), too, adhere to (Possible Worlds), but they believe that possible worlds are mutually isolated regions of spacetime. Hence, concretists define modal notions in terms of possible worlds, but appeal to no modal notions in order to define possible worlds. They, then, are able to *reduce* modal notions to wholly non-modal ones and, consequently, they succeed in answering, without circularity, the question of what the meanings are of modal statements.

For many, however, this reductive analysis of modal discourse comes at a cost too high. Consider:

1. There could have been blue swans.

(1) is intuitively true. If (Possible Worlds) is right, there is a possible world w such that there are blue swans in w. If possible worlds are mutually isolated regions of spacetime, then there is a region of spacetime that is wholly distinct from ours. In the concretist analysis of modal discourse, then, the seemingly sensible conviction that (1) is true entails a troubling ontological commitment: if (1) says that there is a possible world where blue swans exist, why are we convinced that (1) is true? How did we learn about the existence of this world if it is located outside of the region of spacetime that we inhabit? And why did we end up committed to the existence of a whole world that is distinct from ours by merely countenancing the possibility of there being blue swans? Moreover, our concretist ontology will swell monstrously when we realise that there is a

---

74. See Menzel (2021) for an overview of the debate about what possible worlds are.

region of spacetime for each and every one of the states of affairs that we could regard as possible: one, for example, where swans are not blue, but red; one where there are blue swans and red ones as well; one where it is not swans but swallows that are red and blue.

Adherence to the principle of parsimony, then, and the refusal to generate obnoxious epistemological problems may reasonably dissuade many to adopt the concretist analysis of modal discourse despite it possessing the virtue of being reductive. Nevertheless, strong modal fictionalists (henceforth simply 'fictionalists') believe that there is a way to rip off the benefits of this reductive analysis without paying its ontological price. Let us see.

2 Fictionalism

Lewis's concretist theory (1986) states that there are infinitely many, maximally connected regions of spacetime, and that together they constitute logical space. One of these regions is the one that we inhabit, and we call it 'the actual world'. On Lewis's view, though, actuality is not a special kind of existence, but a mere relative property: a world is actual only relative to certain individuals; and, more specifically: for any individual a within logical space, the actual world is the world that a inhabits. All possible worlds, then, *exist* in precisely the same sense as our world.

For Lewis, this theory is a truthful description of reality. For Rosen (1990), by contrast, it is a mere fiction: the fiction of possible worlds. On Rosen's view, this fiction is a very special one, for modal sentences are to be analysed as statements about its contents:

Rosen

For every proposition p:
  i. 'Possibly p' is true ↔ according to the fiction of possible worlds, there is a possible world w such that, at w, p.
  ii. 'Necessarily p' is true ↔ according to the fiction of possible worlds, every possible world w is such that, at w, p.

(The semantic ascent in (Rosen) serves a purpose that will become apparent in the next section.) Rosen claims that (Rosen) equates Lewis's concretism in explicative power: it reduces modal notions to non-modal ones, and constitutes, therefore, a non-circular response to the question of what modal statements mean. Nevertheless, Rosen

maintains, the endorsement of (Rosen) is ontologically much more inexpensive than concretism. According to (Rosen), if (1) is true, then (2) follows:

2.   According to the fiction of possible worlds, there is a possible world w such that there are blue swans in w.

Compare (2) with (3):

3.   According to Wells's *The War of the Worlds*, there are dark-eyed, tentacled aliens invading the Earth.

Clearly, (3) does not commit us to the existence of dark-eyed, tentacled aliens invading the Earth. (2), by parity of reasoning, does not commit us to possible worlds. Consequently, the truth of (1) involves no unwanted ontological commitments.

Rosen maintains, moreover, that (Rosen) surpasses the concretist analysis of modal discourse not only because it is ontologically inexpensive, but also because it does not give raise to any epistemological puzzles. According to Lewis, the existence of worlds in logical space is governed by *the principle of plenitude*, which states, roughly, that, for every recombination of actual objects or parts thereof, there is a possible world where such recombination exists (1986, pp. 86–92).[75] Accordingly, for Rosen, the fiction of possible worlds affirms the existence of the worlds that are generated by the principle of plenitude. And, as it happens, Rosen claims, our imaginative constructions of possibilities obey this principle: the scenarios that we conceive of as possible are precisely those which consist of recombinations of actual objects or their parts. Hence, Rosen contends, when we engage in imaginative experiments, we discover precisely that which is true according to the fiction of possible worlds. Engaging in imaginative experiments, then, results in the acquisition of knowledge about the content of the fiction; and, given that modal claims are ultimately claims about the content of this fiction, our imaginative activities may be rightly identified as the source of modal knowledge.

Multiple objections have been advanced against Rosen's fictionalist theory, and they have all been responded to with plausibility.[76] I think, however, that Rosen's theory

---

75. Strictly speaking, what the principle of plenitude says is that, for any recombination of *counterparts* of objects or their parts, there is a possible world where such recombination exists. I will not discuss the counterpart relation in this paper, however. See Lewis (1971; 1973; 1978; 1986, pp. 192–263).

76. The most prominent objections to Rosen's theory are the Brock-Rosen objection (Brock 1993; Rosen 1993) and Hale's dilemma (1995). Menzies and Pettit (1994) and Noonan (1994), among others,

suffers from a problem that has not been payed enough attention yet.

3   The Lewisian Fiction

What does the fiction of possible worlds say? It says, of course, that there is a plurality of worlds—but which? Lewis never answers this question exhaustively. As I mentioned in the previous section, Lewis believes that logical space is governed by the principle of plenitude; but he himself admits that there are more possible worlds than those produced by this principle: there are, for instance, possible worlds inhabited by *alien individuals*—i.e., individuals that are not recombinations of actual objects nor parts thereof (1986, pp. 91–92). Lewis, in sum, illustrates the spectacular richness of logical space, but does not ever wholly determine its extension.

For Lewis, this is unproblematic. He is a realist: for him, there is a fact of the matter about what worlds exist, even if he cannot tell us exactly which worlds these are. But matters are different for Rosen: for Rosen, Lewis's theory is just a fiction; and, if it does not tell us exactly what worlds exist, then it is *incomplete:* it is entirely silent about some of the features of logical space.[77] Let n be a very large natural number, and consider:

4.   There could have been n non-overlapping physical objects.

According to (*i*) in (Rosen), (4) is true just in case:

5.   According to the fiction of possible worlds, there is a possible world w such that there are n non-overlapping physical objects in w.

Is there, in the Lewisian fiction, a possible world in which there exist n non-overlapping physical objects? Lewis (1986, pp. 101–104) maintains that there is an upper bound m to the number of non-overlapping physical objects that inhabit the most populous universes, but he does not tell us what it is—in this respect, too, the Lewisian fiction is incomplete. The Lewisian fiction, then, does not say that $n \leq m$. (5), consequently, is false. By (Rosen), it follows that (4) is not true—but is it, too, false? Rosen (1990,

---

responded to the former on behalf of Rosen; Rosen (1995) and Divers (1999) replied to the latter. Nolan (2022) offers a detailed reconstruction of these and other objections advanced against Rosen's fictionalist theory, and an overview of their rejoinders.

77. Several authors believe that the incompleteness of the fiction of possible worlds is problematic. See, e.g., Rosen (1990, pp. 341–345); Nolan (1997); Woodward (2012); Skiba (2017).

pp. 341–344) maintains that it is not: that, on pain of absurdity, the modal fictionalist must reject the law of bivalence and admit truth-value gaps.

For the sake of contradiction, suppose that the modal fictionalist adheres to the law of bivalence. Since (4), on their view, is not true, then it is false. Now witness:

6.   Necessarily, there are not n non-overlapping physical objects.

According to (ii) in (Rosen), (6) is true just in case:

7.   According to the fiction of possible worlds, every possible world w is such that there are not n non-overlapping physical objects in w.

As I said, the Lewisian fiction says that there is an upper bound m to the number of non-overlapping physical objects that inhabit the most populous universes, but it does not tell us what it is. It does not say, therefore, that $n > m$. It follows that (7) is false; and, by (ii) in (W'), (6) is not true. If fictionalism abides by the law of bivalence, then (6) is false—but now (6) and (4) have the same truth value despite being contradictories.

On Rosen's eyes, in order to avoid this result that he deems to be absurd, the fictionalist ought to refuse the law of bivalence, and supplement (i) and (ii) in (Rosen) with two clauses meant to prevent that contradictory modal statements receive the same truth value within the fictionalist framework (1990, pp. 341–344):

iii.   'Possibly p' is false $\leftrightarrow$ according to the fiction of possible worlds, it is not the case that there is a possible world w such that, at w, p.
iv.   'Necessarily p' is false $\leftrightarrow$ according to the fiction of possible worlds, it is not the case that every possible world w is such that, at w, p.

By clause (iii) in (Rosen), (4) is false just in case the Lewisian fiction says that there is no possible world where n non-overlapping objects exist. The fiction does not say this—so (4) is not false. Hence, it is neither true nor false: it is a truth-value gap. But (4), of course, is merely an arbitrary instance of the incompleteness of the Lewisian fiction and, thus, our result generalises (Rosen, 1990, p. 343):

Result 1
For every proposition p, if the Lewisian fiction does not say whether p holds in some possible world or not, 'possibly p' is a truth-value gap.

Now that we have introduced the four clauses of (Rosen), we can discuss the reasons why Rosen's proposal is problematic.

On Rosen's view, we have seen, (4) is a truth-value gap: (Result 1) entails that it is so. Arguably, fictions are abstract linguistic structures that exist whether or not they have been tokened. In Platonic heaven, then, there must exist two fictions, F' and F'', characterised as follows: F' resembles the Lewisian fiction exactly, except because it states that there is a possible world inhabited by n non-overlapping physical objects. F'', too, resembles the Lewisian fiction exactly, but it says that there is no possible world inhabited by n non-overlapping physical objects.

Clearly, both F' and F'' are fictions about alternative universes. Why does Rosen claim that the fiction of possible worlds that we insistently speak about via our modal assertions is the Lewisian fiction and not F' nor F'''?

The fictionalist, it seems, has no principled reason to maintain that the fiction that our modal statements are about is the Lewisian fiction, and not any one of the other fictions that exist in Platonic heaven. In the literature, this charge is known as the charge of *fetishism*. It was originally advanced by Rosen (1990, pp. 349–354), but its paradigmatic formulation is Peacocke's:

> The charge of fetishism lodged against the modal fictionalist is that there is no saying what is so special about the theory of possible worlds that is mentioned in the schema asserted by the modal fictionalist. Since the fictionalist does not take [Lewis's theory] to be true, why should we be peculiarly interested in what follows from it, rather than some other theory? (1999, p. 154).

There are several ways in which the fictionalist could try to respond to this charge, but I believe they are all inadequate.

The fictionalist may want to say that the fiction that our modal statements are about is the one fiction that complies with the truth of modal propositions, and that, as it happens, this fiction is the Lewisian fiction: it says that there is a possible world where p just in case possibly p, and that p obtains in all possible worlds just in case it is necessary that p. Nevertheless, it is a well known fact that this hypothesis turns (Rosen) into a hopelessly circular account of modal discourse: according to (Rosen), what modal statements are true depends on what the fiction of possible worlds says; but the hypothesis that we are considering states that the meaning of the expression

'the fiction of possible worlds' depends on what modal statements are true.

Nolan (2022, §4.3) claims that, whilst fictionalist cannot appeal without circularity to modal truths in order to explain why modal statements are about the Lewisian fiction, they can appeal to our *modal convictions*: not to those modal statements that are true, but to those that we believe to be true. On Nolan's view, the fiction that our modal statements are about is the Lewisian fiction because this fiction honours our modal convictions better than any other—better, for instance, than F' and F''. If the fiction referred to in (Rosen) is the Lewisian fiction, then (4) is neither true nor false; but this is not so if the fiction referred to in (Rosen) is F' or F''. If the fiction in question is F', then, by (*i*) in (Rosen), (4) is true; if the fiction is F'' instead, then, by (*iii*) in (Rosen), (4) is false. Now, if n is sufficiently large, then our pre-theoretical intuitions concerning the truth-value of (4) will be completely uncertain: we will not have any convictions as to whether (4) is true or false. Rosen could then argue that the Lewisian fiction, which refuses a truth value to (4), accords with these ambivalent convictions better than do F' and F''. Consequently, Rosen could contend, there is a principled reason to suppose that the fiction that is referred to in (Rosen) is the Lewisian fiction and not F' nor F''. Similar arguments could be used to favour the Lewisian fiction over any one of the other fictions that exist in Platonic heaven.

This response, however, is useless against the fetishism charge. Suppose that, at time t_0, we had a set of modal convictions. Then, at time t_1, we observed the fictions that exist in Platonic heaven, and noticed that one of them complies with our modal convictions better than all the others: it says that there is a plurality of worlds distinct from our own; for every one of the states of affairs that we believe to be possible, there is a world where it obtains, and the states of affairs that we believe to be necessary obtain in all worlds. Let us call this fiction 'F'. Eventually we came to wish to speak about the contents of this very special fiction, and did so by means of modal statements. But what are the modal convictions that we had at time t_0? If belief is a relation that holds between a subject and a proposition, what are the propositions that we believed to be true at t_0? Fictionalists maintain that modal statements express propositions of the following form:

    10.   According to the fiction of possible worlds, p.

When we believed propositions like these at t_0, what was the denotation of 'the fiction of possible worlds'? If it was F, then the modal convictions that we had at t_0 concerned

F itself. Nevertheless, if our modal convictions at time t_0 concerned F, then it is not at all a surprise that F complied with them. Moreover, why did we have convictions about F at t_0 if, by assumption, at t_0 we had not yet realised that F complies with our modal convictions? What was so special about F at t_0? Perhaps nothing was special about F; perhaps we had convictions concerning the contents of several of the fictions that exist in Platonic heaven. If this is the case, though, then the fetishism charge reiterates: why is it that we use modal statements to convey our convictions concerning F and not some other fiction?

An alternative response to the fetishism charge may be drawn directly from Rosen's own work. As I said in the previous section, Rosen thinks that our imaginative activities abide by the principle of plenitude and the other postulates that govern the existence of possible worlds in the Lewisian fiction. Consequently, Rosen maintains, the Lewisian fiction constitutes a systematisation of our imaginative capacities—and this is the reason why we are interested in speaking about it and not the other fictions. Rosen himself recognises that this response to the fetishism charge would create some "affinities" between fictionalism and a theory that he calls 'conceptualism': the theory, he says, that "[aims] to locate the source of modal distinctions in us, in our capacity to imagine or conceive alternatives to the actual state of things" (1990, p. 353).

I believe that the endorsement of the conceptualist contention that the source of our modal convictions are our imaginative capacities is severely pernicious for the fictionalist. Consider:

11.   Possibly p_1.
12.   There is a possible world w such that, at w, p_1.

The fictionalist-conceptualist contention, I take it, says something along the lines of the following: the meaning of a modal statement (11) is the proposition that (12) is true in the fiction of possible worlds. This fiction is an accurate representation of our imaginative capacities. It says that there is a vast multiplicity of worlds distinct from our own and, for every one of the states of affairs that we may conceive, there is a world where it obtains. Consequently, if (12) is true according to the fiction, it is because we can conceive that p_1. If this is the case, though, then an assertion of (11) conveys that, according to an accurate representation of our imaginative capacities, it is conceivable that p_1. Hence, ultimately, an assertion of (11) is nothing more than a roundabout way of affirming that it is conceivable that p_1. But why would we speak in this roundabout

way? Why not ditch the fiction of possible worlds and suppose that the meanings of modal statements are given by the following schema?

Conceptualism

For every proposition p:

 i. Possibly $p \leftrightarrow$ it is conceivable that p.

 ii. Necessarily $p \leftrightarrow$ it is not conceivable that not p.

 iii. 'Possibly p' is a truth-value gap $\leftrightarrow$ it is unclear whether p is conceivable or not.

(Conceptualism) allows us to obtain, without any roundabouts, the same result as before: by (i) in (Conceptualism), an assertion of (11) is an affirmation that it is conceivable that a. The fictionalist, however, could protest that (Conceptualism) is inferior to Rosen's conceptualist fictionalism. Clearly, endorsing (Conceptualism) implies giving up the thesis that modal statements express propositions about the contents of the fiction of possible worlds; and this, Rosen could contend, is undesirable.

But why would it be undesirable? What is the theoretical work supposedly done by the thesis that modal statements express propositions about the fiction of possible worlds? None, it seems to me. If modal statements concern the contents of a fiction, then the question arises why we care about the truth of modal statements as we do.[78] Rosen suggested that the fiction of possible worlds constitutes an accurate account of our imaginative capacities, and he did so in order to explain why we care about the contents of this fiction as we do—more, that is to say, than we care about the contents of virtually any other fictional tale. But then that which explains our relation to modal statements is that they represent our imaginative capacities accurately; no work whatsoever is being done by the thesis that the representation in question is done by way of an appeal to a peculiar fiction about alternative universes. Endorsing (Conceptualism) instead of Rosen's conceptualist fictionalism, then, amounts not to giving up an explanatorily relevant thesis, but to adhering to a more parsimonious conceptualist account of modal discourse—one that does not comprise explanatorily inert tenets. Admittedly, conceptualist analyses of modal discourse like (Conceptualism) are well-known to be problematic; but Rosen himself recognises that fictionalists will face the very same problems if they adhere to his conceptualist version of fictionalism

---

78. This question relates not only to the fetishism problem, but also to the so-called 'argument for concern' against modal fictionalism. See Rosen (1990, pp. 149–154).

(1990, p. 353). There seems to be no reason, therefore, to prefer Rosen's conceptualist fictionalism instead of (Conceptualism).

The fictionalist could protest, however, that (Conceptualism) is not as successful a reduction of modal discourse as (Rosen) because the truth conditions that (Conceptualism) assigns to modal statements are not as crisp as those assigned by (Rosen): checking whether a statement is true according to the fiction of possible worlds, she could contend, is much more straightforward than investigating whether the truth of a given statement is conceivable or not.

To my mind, though, this complaint is misplaced. There is a vast wealth of fictions in Platonic heaven, and they constitute a multi-dimensional, gapless spectrum where each fiction is only ever so slightly different from the ones that are adjacent to it. For instance: the Lewisian fiction is silent with regard to the question of the size of the largest worlds. The Lewisian fiction, moreover, is slightly different from a fiction that resembles it exactly, except because it says that the upper bound to the number of non-overlapping physical objects that exist in the most populous universes is n. This fiction, in turn, is slightly different from a fiction that resembles it exactly, except because it says that the relevant upper bound is $n+1$; etc. If the fictionalist thinks that, on the face of this tremendously rich spectrum, we can determine that one of these fictions specifically complies better than all the others with our imaginative capacities, then surely we have a very precise understanding of the limits of these capacities. And, if this is the case, then verifying whether a given statement p falls within this limit or not should be pretty straightforward. Consequently, the assignment of truth conditions carried out by (Conceptualism) is not theoretically inferior to that carried out by (Rosen).

If this is the case, then the availability of (Conceptualism) implies that fictionalism is an unduly complicated theory of the meaning of modal assertions and, consequently, that, for the sake of parsimony, it ought to be discarded in favour of its simpler alternative—viz. (Conceptualism). Rosen, then, cannot rationally endorse conceptualism and hold on to his fictionalist theory. If he wishes to hold on to his fictionalist theory, that is to say, he must give up his conceptualist response to the fetishism charge, and search for a different one.[79]

---

79. At this point I would like to remark, for the sake of clarity, that my present aim is *not* to defend (Conceptualism). The contention that I am advancing, that is to say, is not at all that (Conceptualism) constitutes the right analysis of modal discourse—or even the best one available. My present contention is simply that conceptualism is preferable to fictionalism *under the assumption* that the fictionalist will

The third response that Rosen could give to the fetishism charge is the one recommended by Woodward (2011), and it consists, in short, in refusing to answer the question of what is special about the Lewisian fiction. Rosen, Woodward argues, need not maintain that modal statements are about *one* fiction that is special amongst all the others: he may contend, instead, that there is a *range* of special fictions that our modal statements are about. Thus Woodward proposes that fictionalists endorse (Woodward):

Woodward

'Possibly p' is true $\leftrightarrow$ according to φ, there is a possible world w such that, at w, p.

According to Woodward, fictionalists should take 'φ' to range over our class of special fictions, and supervaluate (2017, p. 539). The details of this proposal are complex, but I will not reconstruct them here because they are presently unimportant. Let us, then, simply assume that Woodward's supervaluationist strategy results in a reasonable assignment of truth values to our modal statements. What I wish to focus on now are two questions that arise in connection with Woodward's proposal: how is the range of 'φ' defined? And what makes its members so special? Without an answer to the former question, (Woodward) and any other schematic equivalences that the supervaluationist fictionalist may wish to endorse will contain a semantic gap. Without an answer to the latter, the fetishism charge will reiterate.

The question concerning the range of 'φ' could be answered along conceptualist lines. The defender of this approach contends that there is a principled way to interpret the various fictions about possible worlds that exist in Platonic heaven as competing accounts of the limits of our imaginative capacities:

Imaginative Capacities

For every fiction F about possible worlds and for every proposition p:

    i. F says that there is a possible world where $p \leftrightarrow F$ holds, *qua* account of the limits of our imaginative capacities, that it is conceivable that p.

    ii. F says that there is no possible world where $p \leftrightarrow F$ holds, *qua* account

---

follow Rosen in contending that the reason why the Lewisian fiction is special is that it constitutes a systematisation of our imaginative capacities. Since (Conceptualism) is preferable, then, if the fictionalist follows Rosen, she will be compelled by reason to abandon fictionalism and endorse (Conceptualism). Accordingly, if the fictionalist wishes to hold on to her fictionalist theory, she ought to reject Rosen's contention about the reason why the Lewisian fiction is special.

of the limits our imaginative capacities, that it is inconceivable that p.

Amongst the wealth of fictions about possible worlds that exist in Platonic heaven, however, none may be thought to represent the limits our imaginative capacities better than all the others—perhaps these fictions, *qua* accounts of the limits in question, are underdetermined by evidence. Consequently, the limits of our imaginative capacities are best accounted for not by a single one of these fictions but by a range thereof.

Unfortunately, this conceptualist theory that I have just sketched gives way to an argument for parsimony against fictionalism similar to the one that I have formulated already. Let us assume, for the sake of the argument, that this theory is true, and recall (11).

A1   Let $\mathcal{F}$ be the class of fictions that fall within the range of 'φ'.

C1   (11) is true iff φ says that there is a possible world where a—by (Woodward).

C2   (11) is true iff every fiction $F \in \mathcal{F}$ says that there is a possible world where p_1—by (A1) and (C1).

C3   For every fiction $F \in \mathcal{F}$, F says that there is a possible world where p_1 iff F holds, *qua* account of the limits if our imaginative capacities, that it is conceivable that p_1—by universal instantiation on (i) in (Imaginative Capacities).

C4   Every fiction $F \in \mathcal{F}$ says that there is a possible world where p_1 just in case every fiction $F \in \mathcal{F}$ holds, *qua* account of the limits our imaginative capacities, that it is conceivable that p_1—trivially from (C3).

C5   (11) is true iff every fiction $F \in \mathcal{F}$ holds, *qua* account of the limits of our imaginative capacities, that it is conceivable that p_1—by the transitivity of equivalence, (C2) and (C4).

By (C5), an assertion of (11) conveys that all of the members of $\mathcal{F}$ represent it as conceivable that p_1. By assumption, the members of $\mathcal{F}$ are the best representations that there are of the limits of our imaginative capacities. Consequently, an assertion of (11) is, ultimately, nothing more than a tortuous way of conveying that, according to the best representations that there are of the limits of our imaginative capacities, it is conceivable that p_1. But there is a much simpler way to maintain that modal statements like (11) are ultimately about conceivability.

Not only fictions exist in Platonic heaven as abstract linguistic structures: theories

do, too. All kinds of theories about the limits of our imaginative capacities, then, exist in platonic heaven. Moreover, there is a function g that maps $\mathcal{F}$ onto the set of theories about the limits our imaginative capacities in the following way: for any fiction $F \in \mathcal{F}$, let P_F be the set of propositions that, according to F, are true in some possible world, and let P_F' be the set of propositions that, according to F, are not true in any possible world. $g(F)$ is the theory about our imaginative capacities that says that all and only the members of P_F are conceivable, and all and only the members of P_F' are inconceivable. Hence:

Result 2

For any proposition p and any fiction $F \in \mathcal{F}$, F says that there is a possible world where p just in case $g(F)$ says that p is conceivable.

(Result 2) says, in simple terms, that $g(\mathcal{F})$—i.e., the image of g— *mirrors* $\mathcal{F}$: there are theories in $g(\mathcal{F})$ where p is conceivable exactly as there are fictions in $\mathcal{F}$ where there are possible words in which p obtains. By (Imaginative Capacities), if a fiction $F \in \mathcal{F}$ says that there is a possible world in which p obtains, then F, *qua* account of our imaginative capacities, holds p as conceivable. Consequently, for every proposition p and every fiction $F \in \mathcal{F}$, F *says* the same, so to speak, about the conceivability of p as $g(F)$; and $\mathcal{F}$, as a collective body, says the same about the conceivability of p as $g(\mathcal{F})$. More formally:

C1 F_1 says that there is a possible world where p_1 iff F_1 holds, *qua* account of the limits of our imaginative capacities, that it is conceivable that p_1—by universal instantiation on (i) in (Imaginative Capacities).

C2 F_1 says that there is a possible world where p_1 iff $g(F_1)$ says that p_1 is conceivable—by universal instantiation on (Result 2).

C3 $g(F_1)$ says that p_1 is conceivable iff F_1 says that there is a possible world where p_1—commutation on (C2).

C4 $g(F_1)$ says that p_1 is conceivable iff F_1 holds, *qua* account of the limits of our imaginative capacities, that it is conceivable that p_1— by the transitivity of equivalence, (C3) and (C1).

C5 For every $F \in \mathcal{F}$, $g(F)$ says that p is conceivable just in case F holds, *qua* account of the limits of our imaginative capacities, that it is conceivable that p—universal generalisation on (C4).

As I stated earlier: $\mathcal{F}$ and $g(\mathcal{F})$ say the same about the conceivability of any given

177

proposition. Importantly, though, $g(\mathcal{F})$, and each one of its members individually, speak literally about conceivability, whilst $\mathcal{F}$ and its members do so via a detour through the language of the fiction of possible worlds.

If modal statements are ultimately about conceivability, then why should we think of them as being about the contents of the members of $\mathcal{F}$ instead of $g(\mathcal{F})$ if the latter, as I have contended, speak not tortuously but straightforwardly about conceivability? Consider:

Supervaluationist Conceptualism
Possibly $p \leftrightarrow$ according to ψ, it is conceivable that p.

Let 'ψ' range over the elements of $g(\mathcal{F})$. Why should we adhere to (Woodward) instead of (Supervaluationist Conceptualism)? Why should we maintain that modal statements concern the contents of queer fictions that represent the limits of our imaginative capacities in the language of possible worlds if we may just as well claim that these statements concern the limits of our imaginative capacities themselves? Without an answer to these questions, the most parsimonious decision that we can make consists in ditching entirely the analysis of modal statements in terms of fictions about possible worlds, and endorsing (Supervaluationist Conceptualism). A number of worries may be raised concerning the adoption of a conceptualist analysis of modal discourse in place of a fictionalist one; but I think I addressed them all earlier, when I discussed (Conceptualism) and Rosen's response to the fetishism charge.

The result that (Supervaluationist Conceptualism) is preferable to (Woodward) is wholly undesirable for the fictionalist, so let us try to define the range of 'φ' by means other than conceptualism. Woodward's own suggestion is that the fictions that belong to this range are those that honour our modal convictions. Earlier I pondered the hypothesis that the fiction of possible worlds referred to in (Rosen) is the one that honours our modal convictions, and rejected it. I believe that Woodward's proposal concerning the range of 'φ' may be rejected on exactly analogous grounds. Suppose that, at time t_0, we had a set of modal convictions. At t_1, we observed the fictions that exist in Platonic heaven, and realised that some of them comply with our modal convictions better than the others. Let $\mathcal{F}$ be the class of these fictions. It is the elements of $\mathcal{F}$, Woodward says, that 'φ' in (Woodward) ranges over. But what were the modal convictions that we had at t_0? What were they about? (Woodward) says that modal statements express propositions of the following form:

13.   According to φ, p.

When we believed propositions like these at t_0, what was the semantics of 'φ'? If it ranged over the elements of $\mathcal{F}$, then the modal convictions that we had at t_0 were about the elements of $\mathcal{F}$. If this is the case, then it is not at all surprising that the elements of $\mathcal{F}$ complied with our convictions. Additionally, why did we have convictions about the elements of $\mathcal{F}$ at t_0 if, by assumption, at t_0 we had not yet found that the elements of $\mathcal{F}$ comply with our modal convictions better than the other fictions in Platonic heaven? What was so special about the elements of $\mathcal{F}$ at t_0? Perhaps nothing was special about them; perhaps we had convictions concerning the contents of other fictions as well. If this is the case, however, then the fetishism charge resurges: why is it that we use modal statements to express our convictions concerning the elements of $\mathcal{F}$ and not other fictions?

A third alternative for defining the range of 'φ' consists in saying that the fictions within this range are those that adhere to the principle of plenitude and the other principles that govern the existence of possible worlds in Lewis's theory. But what makes these fictions so special? Why do we care so much about their contents and not the contents of other fictions? The fetishism charge has reiterated once again.

4   Another Fiction

In the previous section, I argued that no adequate response exists yet to the question of what fiction is the fictionalists' fiction of possible worlds: neither Rosen, nor Nolan, nor Woodward are capable of telling us what fiction is *the* fiction of possible worlds without circularity, without defeating their own fictionalist contentions, or without prompting the fetishism charge.

Before the question of what the fiction of possible worlds is, though, Armstrong (1989) has a response that differs from the ones that we have reviewed already. Like Rosen, Armstrong adheres to (Rosen). Nevertheless, Armstrong believes that the fiction of possible worlds referred to in (Rosen) is not the Lewisian fiction, but a fiction about possible worlds that he himself described. Perhaps, then, he is able to give us an adequate response (non-circular, non-self-defeating, non-fetishistic) to the question of what the fiction of possible worlds is.

Armstrong thinks that there exist simple individuals—individuals, that is to say, that

have no proper parts—, and universals—i.e., properties and relations that may be instantiated more than once (1989, p. 38–39). Individuals and universals, moreover, are brought together in states of affairs. If individuals $x_1...x_n$ instantiate a universal Φ, then the state of affairs that $\Phi x_1...x_n$ exists (1989, p. 40–41).

Armstrong believes, moreover, that there is a fiction called 'The Book of Worlds', according to which several worlds exist that do not exist in reality. For Armstrong, the Book of Worlds is the fiction referred to in (Rosen). What worlds exist according to The Book of Worlds is a *combinatorial* matter: in principle, each combination of extant individuals and a universal into a state of affairs obtains in a world.[80] Later on, Armstrong qualifies this contention in several ways: he admits, for instance, fictions where states of affairs exists involving individuals that do not exist in the actual world—those that Lewis called 'alien individuals'. Importantly, moreover, Armstrong rejects the existence of negative universals in order to rule out worlds where an individual x instantiates both the property of being Φ and the property of being not-Φ and, thus, where Φx and not-Φx both obtain. If such a world existed in The Book of Worlds, then (Rosen) would determine that possibly Φx and not-Φx—but that is intuitively wrong. Unless (Rosen) is to bring about a revision of our modal convictions, the world in question must be excluded from The Book of Worlds.

Our convictions that it is impossible for a contradiction to be true are strong; but equally strong are our convictions that a state of affairs cannot obtain where an individual instantiates more than one determinate of the same determinable. If these convictions are to be honoured, Armstrong must exclude from his fictions all worlds where an individual instantiates multiple determinates of the same determinable. Clearly, rejecting negative universals cannot help him; so, in order to justify this exclusion, Armstrong (1989, pp. 77–84) appeals to 'analytic modalities': modalities grounded in logic and meaning. Suppose, for instance, that an object a instantiates the property of having a mass of 1 kg and the property of having a mass of 2 kg in world w. If an individual x has a mass of 2 kg, then, by mere analysis, it follows that, for every part y of x that is 1 kg, there exist a part z of x that is 1 kg and does not overlap with y. Now, by assumption, a is 1 kg. Moreover, a is a part of a. It follows that there must be a part

---

80. Kim (1986, p. 597) worries that, for Armstrong to obtain all the possible states of affairs that he wants, 'all combinations' must signify 'all *possible* combinations' and, hence, that his theory is not reductive. But I think Kim's worry is misplaced. Perhaps what Armstrong means is that, for every ordered set $\langle \Phi, x_1...x_n \rangle$, where $x_1...x_n$ are individuals and Φ is a universal, the state of affairs that $\Phi x_1...x_n$ obtains in a world.

b of a that is 1 kg and does not overlap with a. By definition, though, any part of an individual x overlaps with x. If b does not overlap with a, then b is not a part of a. Hence, b is a part of a and it is not the case that b is a part of a. This is a contradiction; it is logically false. Consequently, Armstrong concludes, it is not the case that a is 1 kg and 2 kg in w. He contends, moreover, that an analogous *reductio ad absurdum* may be carried out under any supposition that an individual instantiates multiple determinates of the same determinable in one world.

To my mind, Armstrong's argument is enthymematic, and its hidden premise he cannot endorse without giving up the reductive powers of his theory. On his view, it is not the case that a is 1 kg and 2 kg in w because the supposition that it does entails, by analysis, a logical contradiction. However, it is not at all clear why a logical contradiction cannot obtain in w. Armstrong, it appears, assumes surreptitiously that each description of a possible world in The Book of Worlds must be consistent, but consistency is standardly defined in modal terms. Unless Armstrong gives us a non-modal definition of consistency, his theory is not reductive. Alternatively, he could choose to betray our convictions that states of affairs are impossible where an individual instantiates multiple determinates of the same determinable. He would thereby free himself of the task of ensuring that no such state of affairs obtains in any world. If he were to opt for this alternative, though, then the burden would be on him to convince us that fictionalism is preferable to its ontologically weightier rivals even if the latter, but not the former, honour our modal convictions.

5   Conclusion

I posed one question at the beginning of this paper: modal fictionalists believe that modal statements express propositions concerning the contents of *the fiction of possible worlds*—but what fiction is this fiction? I have argued that no answer has been given to this question which serves the purposes of the strong modal fictionalist, who wants to reduce modal notions to non-modal ones. If my arguments are sound, strong modal fictionalists still owe us an answer: what fiction is their fiction of possible worlds? Unless an adequate response is given, it will remain unclear what, on their theory, modal statements are about.

Chapter Nine:
Moral Fictionalism and Moral Revisionism

A moral proposition is a proposition that is true just in case a certain object instantiates a certain moral property. A moral sentence is a sentence that appears to express a moral proposition: under a face-value reading, this sentence says of an object that it instantiates a moral property. Moral anti-realists all believe that there are no true moral propositions: moral properties, on their view, do not exist, and thus nothing instantiates them. With regard to moral sentences, though, various moral anti-realists endorse diverging views. The error theorist, for instance, thinks that moral sentences do indeed express moral propositions. Since none of these propositions is true, asserting a moral sentence constitutes in every case an error. The non-cognitivist, on the other hand, believes that moral sentences express not moral propositions, but non-cognitive attitudes like approval, disapproval and desire. Upholding a moral commitment, the non-cognitivist contends, consists, accordingly, not in believing a certain moral proposition, but in possessing an attitude of the non-cognitive sort.

Non-cognitivists have traditionally accounted for the non-cognitive character of moral commitment by means of a semantic thesis: non-factualism. Non-factualism states that moral sentences have a non-standard semantics: contrary to what their surface form suggests, they express not moral propositions, but non-cognitive attitudes. Joyce (2001) and Kalderon (2005) independently maintain that non-cognitivists have insistingly endorsed non-factualism because they have been systematically oblivious to the distinction between meaning and use: they have omitted the possibility of locating the non-cognitive character of moral commitment at the level not of semantics, but of pragmatics.

If Joyce and Karderon are right, then non-cognitivism is not wedded to non-factualism. Factualist non-cognitivism comprises a characteristically non-cognitivist characterisation of moral commitment: it states that moral commitment is the possession of non-cognitive attitudes and not the possession of moral beliefs. By contrast with its non-factualist counterpart, however, this theory states that moral sentences have a standard semantics: they express moral propositions, as their surface form suggests. The non-cognitive character of moral commitment lies, accordingly, in the pragmatic level: we *use* moral sentences to express non-cognitive attitudes rather than moral propositions.

Both Joyce and Kalderon claim that factualist non-cognitivism is best termed 'moral fictionalism'. I believe that they are right: the theory in question bears the hallmarks of the fictionalist project. A fictionalist theory is a theory that encompasses three doctrines: an anti-realist thesis concerning the members of a kind K; a standard semantic account of the sentences that quantify existentially over Ks, or purport to refer to Ks; and an anti-revisionary policy with regard to said sentences. An instance of each one of these three doctrines is endorsed by the moral fictionalist. She is an anti-realist concerning moral properties; she attributes a standard semantics to moral sentences; and she adheres to an anti-revisionary policy concerning our use of these sentences: if moral sentences are ordinarily used to express not moral propositions, but non-cognitive attitudes, then we may rationally continue using them as we always have—even if they are all untrue. In more succinct terms: the supreme fictionalist goal is that of encompassing anti-realism concerning Ks with anti-revisionism concerning K-discourse—and doing so without attributing a non-standard semantics to K-sentences.[81] This goal is exactly that which the moral fictionalist pursues; but, in this chapter, I will argue that she has failed in attaining it.

In §1 and §2, I will discuss Kalderon's fictionalist theory of morality. In §3 and §4, I will discuss Joyce's importantly different alternative. Throughout these four sections, I will argue that neither Joyce nor Kalderon has succeeded in achieving the fictionalist's anti-revisionist aim—but I will also argue that moral fictionalism, in the versions in which it currently exists, cannot be rationally endorsed.

---

81. In the following section I will advance a hypothesis concerning the reasons why attributing a non-standard semantics to a sphere of discourse may be thought to be generally undesirable. Non-standard semantic theories about moral discourse, however, have tended to take the form of non-factualist non-cognitivist theories, and these theories are well-known to be problematic in their own special way, for they give rise to the Frege-Geach problem.

1   Theoretical Parsimony

Kalderon (2005, pp. 1–51) argues extensively for the thesis that moral commitment consists in the possession of non-cognitive attitudes. As I said earlier, however, both the fictionalist and the traditional non-cognitivist endorse this thesis. Nevertheless, Kalderon maintains that fictionalism is superior: rejecting non-factualism and attributing a standard semantics to moral sentences, Kalderon claims, gives the fictionalist two mutually independent advantages. The first is theoretical parsimony (2005, p. 118); the second, immunity to the well-known Frege-Geach problem (2005, pp. 10–11).[82]

The contention that that fictionalism is immune to the Frege-Geach problem has been robustly denied by Lenman (2008) and Eklund (2009). If their objections are sound, then the only advantage that is left for the fictionalist to claim over the non-cognitivist is that which corresponds to theoretical parsimony. Nevertheless, I will argue in what follows that Kalderon is wrong: fictionalism is not more parsimonious than traditional non-cognitivism.

Both the traditional non-cognitivist and the fictionalist adhere to a non-cognitivist conception of moral commitment. Each one of them, moreover, adheres to a further thesis: the traditional non-cognitivist adheres to non-factualism; the fictionalist, to the thesis that moral sentences are ordinarily used to express non-cognitive attitudes. If this is the case, why does Kalderon claim that fictionalism is more parsimonious?

The answer, I believe, is simple. Non-factualism is the thesis that moral sentences have a non-standard semantics, and the theses that attribute a non-standard semantics to pieces of discourse are theoretically costly: they tend to be markedly counterintuitive, and they fracture the unity of the theories that exist concerning our various discursive spheres. Consequently, dispensing of theses of this kind contributes to theoretical parsimony.

Importantly, however, transforming traditional non-cognitivism into fictionalism does not only consist in expelling non-factualism. In order to explain the non-cognitive character of moral commitment, the fictionalist ought to take on a thesis about the pragmatics of moral discourse, and this thesis, I will now show, is not less of a theoretical burden than non-factualism.

---

82. See Searle (1962) and Geach (1965) for the original formulation of the Frege-Geach problem.

According to Kalderon, the fictionalist endorses the following thesis concerning the pragmatics of moral discourse: moral sentences are not ordinarily asserted, but *quasi-asserted*. 'Quasi-assertion' Kalderon defines negatively as follows: an utterance of a sentence s is a quasi-assertion when it is *not* an assertion of the proposition expressed by s. Ordinary utterances of moral sentences, Kalderon affirms, do not express moral propositions, and the reason why this is so is that these utterances are quasi-assertions, rather than assertions (2005, pp. 119–120).

Kalderon (2005, p. 141) claims that, intuitively, when we utter a moral sentence, we assert a moral proposition—a proposition, that is to say, that a certain object instantiates a certain moral property. Non-factualism contradicts these intuitions: it states that no moral propositions are asserted by way of moral utterances. Nevertheless, Kalderon's own claim implies that the quasi-assertion thesis, too, contradicts our intuitions: our intuitions have it that moral utterances are assertions, and the thesis in question says that they are not. By Kalderon's own lights, therefore, the quasi-assertion thesis is not more benign towards our intuitions than non-factualism.

Moreover, theoretical unity is fractured by the quasi-assertion thesis just as much as it is by non-factualism. Moral utterances are seamlessly interwoven with, and phenomenologically indistinguishable from, referential discourses, *and also* assertoric discourses. Picturing moral utterances as oddities, then, is a distasteful consequence of both non-factualism and Kalderon's quasi-assertion thesis.

To my mind, this should suffice to establish that fictionalism is, *pace* Kalderon, not more parsimonious than traditional non-cognitivism—but there is more to say.

In order to elaborate on his views, Kalderon resorts to a characteristically fictionalist device: the distinction between the real and the fictional content of an utterance (2005, p. 120). When a sentence is used figuratively, the corresponding utterance possesses both a real content and a fictional content. Consider the following piece of figurative speech: "In the flexible mirror of nature, / we are fish, stars are seine, / gods are ghosts in the darkness" (Khlebnikov, 2000, p. 379). Its literal content is its fictional content; what the metaphor *really says*—the thought that it metaphorically conveys—is its real content.

When a sentence s is asserted, Kalderon says, the proposition p expressed by s is the real content of the relevant utterance. When s is quasi-asserted, by contrast, p is s's

185

fictional content. Kalderon, we have seen, states that moral sentences are ordinarily quasi-asserted. Thus, on his view, moral propositions constitute not the real contents of moral utterances, but their fictional contents. Moral sentences, we ought to conclude, are used to convey not moral propositions but something else.

But what is that which moral sentences are used to convey? What is the real content of moral utterances? The answer appears obvious: non-cognitive attitudes. At this point, though, Kalderon makes it clear that this response is not entirely correct. His version of non-cognitivism, it turns out, is atypical: it states that moral acceptance is not wholly non-cognitive, for it consists of "an amalgam between cognitive and non-cognitive attitudes" (2005, p. 129). According to Kalderon, then, the real content of moral utterance is a non-cognitive attitude combined with a cognitive component; and the non-cognitive component is, in turn, a non-moral proposition that describes "the morally salient features of the relevant circumstance" (2005, pp. 129)—the circumstance, that is to say, that the moral utterance concerns.

In sum, then: the fictional content of a moral utterance is a moral proposition; its real content is an amalgam of a non-cognitive attitude and a non-moral proposition. Now we get to put in its place the final piece of Kalderon's theory. The fictional content of a moral utterance, he claims, has a "framing effect" over its real content: the fictional content *frames* the real content so as to represent it under a normative light (2005, p. 149). Equivalently: moral propositions frame our non-cognitive attitudes, and the morally salient features of the relevant circumstance, so as to represent them under a normative light.

On Kalderon's view, therefore, moral propositions are not expressed by ordinary moral utterances, but they do play a role in moral discourse: they infuse our moral utterances with a normative character. When considering, say, torture, we could use an interjection to express our repulse for it. We could also describe its morally salient features by way of a non-moral proposition—we could say, for example, that it produces suffering. By uttering the appropriate moral sentence, however, we do both things at once and even more: when we say that torture is wrong, we present the morally salient features of torture, and our repulse for it, under a normative light.

According to Kalderon, framing powers of the kind that moral propositions possess are possessed by pieces of figurative speech in general. In general, Kalderon maintains, the fictional content of a figure of speech has a framing effect over its real content: the

real content concerns a particular object, and this object is represented under a certain qualitative light by the fictional content of the figure in question (2005, pp. 147–149). Let us consider, for instance, the passage by Khlebnikov that I quoted above, and, more specifically, the following metaphor: "we are fish". On Kalderon's view, this metaphor has a real content, and this real content has an object—humans, presumably. The fictional content of the metaphor has a framing effect over this object: it presents its object under a certain qualitative light—and, in this particular case, the qualitative light, perhaps, of the insignificant.

Kalderon suggests, ultimately, that moral sentences constitute a peculiar fiction: the fiction of morality. And he claims, moreover, that the fiction of morality is very useful. Without it, we could use non-moral sentences to describe the morally salient aspects of the portions of reality that concern us, and we could also express the non-cognitive attitudes that we possess in relation with said portions of reality. The moral fiction, however, allows us to do all of these things with a normative tone. Hence, the moral fiction plays, in our moral practices, the substantive role of instilling normativity in our non-cognitive attitudes and our descriptions of the morally salient features of reality.

There are many things to be said about Kalderon's theory. At this point, however, I wish only to resume my discussion of Kalderon's contention that fictionalism is theoretically more parsimonious than traditional non-cognitivism. Kalderon thinks that fictionalism is superior in its parsimony because the traditional non-cognitivist endorses non-factualism, whilst the fictionalist does not. Non-factualism, I suggested earlier, is theoretically costly because it is counterintuitive, and fractures the unity of the theories that concern our various discursive spheres. A non-cognitivist that does without non-factualism, therefore, endorses a more parsimonious theory than the non-cognitivist who adheres to non-factualism. Nevertheless, the fictionalist cannot be correctly described as a non-cognitivist that does without non-factualism: she does reject non-factualism, but she adopts, in its place, a plurality of theses that are at least as costly. Earlier I argued that Kalderon's quasi-assertion thesis is hostile, as much as non-factualism, to both theoretical unity and our intuitions. Importantly, though, other fictionalist theses are hostile in this way as well. Let us see.

Let T_1 be the fictionalist thesis that moral utterances have both a fictional and a real content; let T_2 be the fictionalist thesis that the fictional content of a moral utterance has a framing effect over its real content. As I mentioned above, Kalderon (2005,

p. 141) maintains that, intuitively, when we utter a moral sentence, we assert a moral proposition. We have seen that the quasi-assertion thesis contradicts these intuitions: it states that moral utterances are not assertions at all. Nevertheless, T_1 and T_2 contradict our intuitions as well. Our intuitions, Kalderon affirms, tell us that what we do by means of a moral utterance is to assert a moral proposition. These intuitions, though, are wrong on Kalderon's own fictionalist theory. According to T_1, moral propositions are the fictional, rather than the real contents of moral utterances. Hence, that which we effectively convey by means of a moral utterance is not a moral proposition at all. According to T_2, moreover, moral propositions are nothing more than framing devices for the real contents of moral utterances. Both T_1 and T_2, therefore, are frankly at odds with Kalderon's own account of our intuitions: both entail, against these intuitions, that what a moral utterance conveys is not a moral proposition.

If my conclusion above is sound, then fictionalism is patently not more parsimonious than traditional non-factualism. Non-factualism may well be a counterintuitive thesis, and it might well cause fractures in the unity of our theories. Nevertheless, fictionalism is not the result of subtracting non-factualism from non-cognitivism. The fictionalist contends that moral sentences are not used to express moral propositions—and yet, that moral propositions constitute a moral fiction that plays a crucial role within our moral practices. In order to advance this contention, the fictionalist adds to non-cognitivism a plurality of doctrines; she adds, in particular, the quasi-assertion thesis, T_1 and T_2. And these theses, I have shown, are not at all benign towards theoretical unity or our intuitions—not even towards Kalderon's own constructions thereof. Consequently, they are not more benign that traditional non-cognitivism. I conclude, therefore, that fictionalism is, *pace* Kalderon, not superior in its parsimony to its traditional rival.

2   The Plurality of Fictions

As said in §1, moral fictionalism is a characteristically fictionalist theory in that it is supposed to encompass three doctrines: anti-realism; a standard account of the semantics of moral sentences; and an anti-revisionary policy concerning our use of moral sentences. Comprising a standard account of moral sentences was supposed to give fictionalism two advantages over traditional non-cognitivism: theoretical parsimony, and immunity against the Frege-Geach problem. As I mentioned earlier, Lenman (2008) and Eklund (2009) have forcefully challenged the thesis that fictionalism is immune to the Frege-Geach problem. Moreover, in the previous section I argued that fictionalism

is not superior in its parsimony to traditional non-cognitivism. If my arguments are sound, then Kalderon has not yet given us a reason to prefer fictionalism over its rival.

In what follows, I will argue, additionally, that Kalderon's version of moral fictionalism is not what it is supposed to be. In particular, I will show that this version of moral fictionalism is not anti-revisionary. At the end of this section, I will show that my argument has catastrophic consequences for the project of moral fictionalism.

I will start my argument by discussing the various ways in which Kalderon motivates his anti-revisionary doctrine. According to Kalderon (2005, p. 142), our moral commitments are embodied in our use of moral discourse. Upholding a certain moral commitment consists in accepting a certain moral sentence, or, equivalently, in being disposed, in certain situations, to put said sentence forward. Thus, on Kalderon's view, anti-revisionism concerning our use of moral sentences, and anti-revisionism concerning our moral commitments, go hand in hand. Kalderon (2005, pp. 141–142) claims, moreover, that his theory does not entail that any revision is required to our moral commitments. Consequently, on his view, no revision is required to our use of moral sentences.

In order to motivate anti-revisionism, Kalderon claims, as we saw earlier, that moral sentences constitute a fiction that plays a crucial role in our moral practices: it allows us to instil normativity in our non-cognitive attitudes, and our descriptions of the morally salient features of reality. Since moral sentences, despite being untrue, play this crucial role in our moral practices, we ought not revise them.

Kalderon maintains, moreover, that what we do with moral discourse is not to try to describe moral reality. Instead, we build a practice of enabling speakers to respond affectively and to act in a certain way. The business of moral discourse, therefore, is seeking moral transformation, not aiming at truth. Consequently, Kalderon affirms, the fact that moral sentences are untrue is unimportant for moral practice, and thus we need not to revise our use thereof (2005, pp. 151–152).

Kalderon's final contention in favour of anti-revisionism goes as follows: even if moral commitment is not belief in moral propositions, there is a sense in which we are *justified* in having the moral commitments that we currently have. This peculiar kind of justification, Kalderon maintains, is internal to the moral practice, rather than justification *simpliciter* (2005, pp. 145–146); but, on Kalderon's view, it appears, this peculiar kind of justification suffices to ensure that we need not revise our moral commitments, nor,

therefore, our use of moral sentences.

Kalderon does not say much more about what this special sort of justification consists in. I believe, though, that one can sensibly guess that it consists in fidelity to the moral fiction. Roughly, I think, a subject a is justified in accepting that an object b has moral property F iff, according to the moral fiction, b is F. My reasoning behind this suggestion is the following: it is the usefulness of the moral fiction in our moral practices, I suppose, what justifies us in accepting moral sentences. If this is the case, then the sentences that we are justified in accepting are those that constitute the moral fiction.

Whether this suggestion is right or not, however, is ultimately irrelevant for my purposes. What I want to show now is that Kalderon's conception of the moral fiction severely undermines his attempts to establish anti-revisionism.

A crucial question to ask in relation to the moral fiction is how its contents are determined. Kalderon put forward a principle that answers this question:

> **Kalderon's Principle**
> According to the moral fiction, $G_1a \leftrightarrow$ in reality, a has a non-moral property G_2, which, in a subject with a virtuous moral sensibility, would elicit the affective response associated with F (2005, pp. 150–151).

A relatively concrete example might be very helpful in understanding what (Kalderon's Principle) amounts to. Suppose that, in a certain situation, φing has the non-moral property of being such that it will cause suffering to a child. In a person with a "virtuous moral sensibility", this non-moral property will elicit the affective response of desiring that nobody φs, so that the child does not suffer. Then, according to the moral fiction, φing is wrong or impermissible.

A consequence of (Kalderon's Priciple), acknowledged by Kalderon (2005, pp. 150–151), is that the content of the moral fiction is ultimately determined by our conception of moral virtue. Suppose that the non-moral properties of the action of φing are such that φing elicits approval or desire in a certain subject a. If we take a to be virtuous, then the moral fiction will say that φing is good or required by duty. Suppose, by contrast, that φing elicits revulsion or disapproval. If we take b to be virtuous, then the moral fiction will say that φing is wrong or impermissible. Ultimately, then, what the moral

fiction says is determined by our decision to identify a certain subject, and not others, as virtuous.

There are, however, many different conceptions of moral virtue, and, according to Kalderon, they generate distinct, and potentially competing moral fictions associated to various diverging moral systems (2005, pp. 150–151). We have seen, however, that the fictionalist is an anti-realist: on her view, there are no moral properties and, hence, there are no moral virtues possessed by subjects. There is no such thing, therefore, as the one true conception of moral virtue. Consequently, none of the moral fictions is superior to the others. Moral justification, moreover, exists only relative to a moral fiction, so there is no moral justification for preferring one fiction over the others. Finally, nowhere in Kalderon's proposal do we find a suggestion that there might be non-moral reasons of some sort to privilege one conception of moral virtue, nor one fiction, over the rest. They are all, it appears, equally good.

At this point it seems that Kalderon is committed to a pluralistic-relativistic view of morality. Multiple moral fictions are equally legitimate, and each one of these fictions produces a unique moral system. Hence, there exists a plurality of equally legitimate moral systems. Moreover, moral propositions are all untrue, for moral properties do not exist. As (Kalderon's Principle) suggests, certain moral propositions are true relative to certain fictions, or, more precisely, they are true in certain fictions—but no moral proposition is true *simpliciter*. There is, therefore, no ultimate moral truth—no moral truth at all beyond the various moral fictions that exist.

To my mind, this pluralistic-relativistic stance towards morality is in various ways incompatible with anti-revisionism. Briefly, my contention is this: if somebody comes to accept Kalderon's theory, and thereby comes to accept that there is a plurality of equally legitimate moral practices, but no ultimate moral truth, she will have very good reasons to revise her moral commitments.

Suppose that a subject a adopts, at time t, Kalderon's pluralist-relativist theory. She then ponders the fact that there is a plurality of equally legitimate moral systems, and resolves that the only sensible stand to take towards them is that of tolerance and non-interventionism. Since the various systems that exist are as legitimate as a's own, a concludes that it would be wrong to interfere with the moral practices of the advocates of other moral systems in the name of her own moral commitments. But let us see how this conclusion fares in practice.

Suppose that a certain moral system S recommends the subjection of women to men. S states, moreover, that, if a woman refuses to abide by a man's will, it is legitimate to exercise violence against her. a's moral system, by contrast, radically forbids violence and the subjection of persons to the will of other persons. At some point, a encounters a woman and a man who adhere to S. She refused to abide by his will, and is being assaulted now as a consequence. The moral commitments that a held prior to t recommend intolerance and interventionism on the face of a situation like the one that she has encountered. Nevertheless, a adhered to Kalderon's theory at t, and concluded, as a consequence, that intolerance and interventionism are wrong. Thus a will now revise her pre-t moral commitments: she will act, on the face of the situation in question, not in accordance with these commitments, but something else.

In more general terms: if one's own moral system recommends intolerance or interventionism in certain situations, revisions will be rationally demanded by Kalderon's theory: this theory states that other moral systems are equally legitimate, so, on this theory, there is no moral nor non-moral justification for intolerance or interventionism.

Alternatively, upon adopting Kalderon's theory, a could resolve that, since there is no ultimate moral truth, there is no reason, neither moral nor non-moral, for her to keep her pre-t moral commitments. She thus abandons these commitments, and executes, as a consequence, the most radical of moral revisions.

There are, in sum, at least two simple and judicious lines of reasoning that would lead a sensible subject to revise her moral commitments after adopting Kalderon's theory. Kalderon, however, claims that his theory does not lead to moral revisionism. If my arguments above are sound, then Kalderon is wrong.

Moral anti-realism states that moral properties do not exist and, accordingly, that nothing instantiates moral properties. At the beginning of this chapter, I said that a moral proposition is a proposition that is true just in case a certain object instantiates a certain moral property. If nothing instantiates moral properties, then the following conclusion follows:

Conclusion
All moral propositions are untrue.

(Conclusion), in sum, is a direct consequence of anti-realism. Now, if upholding a moral

commitment consists in believing a moral proposition to be true, then (Conclusion) implies that all of our moral commitments are wrong. If we assert moral sentences in order to express moral propositions, then, by (Conclusion), our use of moral sentences is irrational. Anti-realism, then, by way of (Conclusion), appears to lead us straight to the result that our moral commitments and our moral discourse are, by the standards of reason, seriously defective, and, consequently, in need of revision. This result may be understandably judged to be undesirable. Revisionism is naturally unappealing: abandoning our convictions is difficult and unpleasant at many levels. Nevertheless, I believe that, in regard to morality, revisionism does not only appear unappealing, but also potentially catastrophic: as Joyce (2001, pp. 180–185) notes, moral commitments and moral discourse constitute crucial, irreplaceable devices for regulating one's own and others' behaviour.

Kalderon adheres to moral anti-realism; and, in order to block the undesirable result that it produces, he carries out significant amounts of theoretical work: he argues that upholding a moral commitment consists not in believing a moral proposition, but in possessing a non-cognitive attitude; he argues that moral sentences are not used to express propositions; that the moral sentences that we accept constitute a fiction that plays a crucial role in our moral practices; that moral discourse promotes moral transformation; and that there is a sense in which we are justified in having the moral commitments that we do. If my arguments above are sound, however, then all of these arguments are entirely futile: the dreadful burden of revisionism came upon Kalderon despite his ponderous efforts to avoid it.

At the beginning of this chapter I said that the supreme fictionalist goal is that of encompassing, for some kind K, anti-realism about Ks with anti-revisionism concerning K-discourse—and to do so without attributing a non-standard semantics to K-sentences. If the arguments that I advanced in this section are sound, then Kalderon has failed in attaining the goal in question. He attributed a standard semantics to moral sentences, and he upheld anti-realism about moral properties—but he failed in encompassing the latter with an anti-revisionary doctrine about moral discourse.

Does this mean that moral fictionalism is unrealisable—that a standard semantics about moral discourse and moral anti-realism are simply inconsistent with moral anti-revisionism? I believe that it is still too soon to draw this conclusion. In (2001), a fictionalist theory of morality that is importantly different from Kalderon's was pro-

posed by Joyce; and, to my mind, this theory seems especially promising with regard to the fictionalist aim of attaining anti-revisionism. In the following section, I will present the theory in question, and explain why it seems thus promising. Then, in §4, I will investigate whether it is only promising, or also effectively successful.

3   Revolutionary fictionalism

Joyce, like Kalderon, is an anti-realist: he believes that there are no moral properties, and, consequently, that there are no moral facts—no fact exists, that is to say, of an individual instantiating a moral property. And yet, Joyce rejects non-cognitivism *tout court*—both the non-factualist and the fictionalist strands. He thinks, therefore, that moral commitment is moral belief; that moral sentences express moral propositions; and that we use moral sentences to express moral propositions (2001, pp. 10–13). Finally, since we believe moral propositions and assert moral sentences, we are in error.

Joyce, in sum, advocates for an error theory of morality.[83]   And, according to him, there are three alternative paths we may follow in response to moral error. The first is eliminativism: we may abandon moral discourse entirely. The second is self-deception: we may retain our current moral commitments even though we know that they are wrong. The third one is fictionalism: we may keep advancing moral sentences as we always have—but do so without believing the moral propositions that they express.

For Joyce, the first two alternatives are hateful: self-deception is irrational (2001, p. 178), and eliminativism requires us to renounce a crucial means to regulate our own behaviour (2001, p. 181). Fictionalism, by contrast, allows us to preserve moral discourse and, hence, this crucial means to regulate our behaviour. Moreover, fictionalism permits us to do this without forcing us to incur self-deception, nor, therefore, irrationality, for the fictionalist suggests that we use moral sentences without believing the untrue moral propositions that they express.

The fictionalist theory that Joyce has in mind is quite similar to Kalderon's: it comprises a non-cognitive conception of moral commitment, and a thesis about the pragmatics of moral discourse. Nevertheless, the attitude that Joyce holds towards this theory is substantively different from Kalderon's: Kalderon believes that the theory in question

---

83. Joyce (2001, pp. 1–52) argues for his error theory extensively, but, due to considerations of space, I shall not reconstruct his arguments here.

is true; Joyce, by contrast, believes that it is currently false, but we should make it true in response to the moral error. Whilst Kalderon's fictionalism is *hermeneutical*, therefore, Joyce's is *revolutionary*.

Let us now look into the details of the fictionalist theory that Joyce suggests that we adopt. On his view, we should *accept* moral sentences instead of believing the propositions that they express. A subject a accepts a sentence s iff a is disposed to assent to s in a number of situations, but will deny s in *critical* contexts—contexts, that is to say, where she is rigorously inquired about truth and her beliefs. Moreover, if a accepts s, then she is *making a fiction of s*. Thus Joyce suggests that we make a fiction of morality (2001, pp. 193–196). Importantly, Joyce claims, accepting a sentence s that we know to be untrue is not a form of self-deception: the proof that it is not is given by the fact that we will deny s in critical contexts.

Let us momentarily grant Joyce that acceptance is not a form of self-deception. A question suggests itself at this point: if there are no moral facts, why should we accept moral sentences? Joyce says that we do not want the kind of unregulated behaviour that would arise if we abandoned moral discourse entirely, but why not? The results of acting in accordance with morality, Joyce says, are desirable. Of course, however, he does not think that they are morally desirable: he thinks, instead, that they are desirable in a merely pragmatic or instrumental sense (2001, pp. 181–185).

Joyce's contention, then, is that the consequences of behaving morally are desirable independently of the fact that morality is grounded on the erroneous assumption that there are moral facts. Hence, the ends of morality are to be pursued even if we know that morality is an error.

At this point it seems unclear why Joyce thinks that fictionalism surpasses eliminativism. Joyce says that we should adopt fictionalism and keep using moral discourse because the ends this discourse guides us towards are instrumentally desirable. Hence, he seems to think that now that we have learned that there are no moral facts, that which is left for us to pursue is the instrumentally desirable. Surely, though, the eliminativist pursues the instrumentally desirable, too. She might have abandoned morality, but she has not abandoned rationality. If she is rational, and she endorses no moral restrictions of any kind for her own behaviour, then she will surely pursue that which is instrumentally best for her. If this is right, then Joyce is wrong: there is no reason to prefer fictionalism over eliminativism.

In order to block this result, Joyce (2001, pp. 212–213) claims that the fictionalist has a powerful tool that the eliminativist lacks in their common pursuit of the instrumentally desirable. The tool that the eliminativist possesses to instigate herself towards the instrumentally desirable is the hypothetical, instrumental imperative. Nevertheless, Joyce claims, we are all well aware that instrumental imperatives are frequently ineffective: weak will often gets in the way of our goals—regardless of how persistently we give ourselves the relevant instrumental imperatives. The fictionalist, by contrast, possesses the categorical imperatives of morality. These imperatives, Joyce maintains, will guide the fictionalist towards the instrumentally desirable, but they will do so with the special "must-be-doneness" that only imperatives of this kind possess. Thus pursuing the instrumentally desirable with the aid of categorical imperatives makes the fictionalist less likely than the eliminativist to suffer from *akrasia*.

Joyce's contention, in sum, is that the categorical imperatives of morality, which only the fictionalist possesses, make for better behaviour-regulating devices than the instrumental imperatives that the eliminativist is left with after abandoning morality. Let us thus grant, for now, that Joyce is right: fictionalism is preferable.

A summary of Joyce's proposal at this point goes as follows: all the moral sentences that we believe we should stop believing because they are untrue—but we should not abandon them altogether: we should *accept* these sentences, thereby making a fiction of them. In absence of moral facts, what is left for us to pursue is the instrumentally desirable, and moral sentences will guide us towards it whilst guarding us effectively from *akrasia*. The fiction of morality, in other words, will help us 'rehearse the thought' that certain things ought to be done categorically because they are good, and other things ought to be avoided categorically because they are wrong. And this, in turn, will help us attain and avoid, respectively, precisely those things that we would want to attain and avoid for wholly non-moral reasons. In the remainder of this section, I will show why Joyce's proposal is relevantly different from Kalderon's.

Joyce, we have seen, claims that the ends that moral sentences guide us towards coincide with the instrumentally desirable. To my mind, it is clear that, in making this claim, Joyce does not refer to just *any* moral sentences. Suppose that two competing moral systems yield a pair of conflicting moral sentences: one of them recommends φing, and the other one forbids it. If Joyce said that *all* moral sentences, pertaining to all moral systems, lead to instrumentally desirable ends, then his theory would imply that both

φing and not φing lead to the instrumentally desirable. That both φing and not φing both lead to the instrumentally desirable is not necessarily absurd: surely there are cases where both performing and action and not performing it deliver instrumentally desirable results. Nevertheless, that both φing and not φing lead to the instrumentally desirable is not true for every φ. For instance: Joyce (2001, pp. 180–182) insists that cooperation between peoples is instrumentally better than murdering and enslaving. He seems to think that cooperation produces greater benefits, and prevents enmity and subsequent violence. If this is the case, then Joyce cannot then consistently say that a moral system that does not recommend cooperation is equally efficient as a system that does at leading us towards the instrumentally desirable. If Joyce is right that cooperation is, for purely instrumental reasons, more desirable than murdering and enslaving, then it is only one of the systems that we are considering that leads us towards the instrumentally desirable.

I thus conclude that, when Joyce says that moral sentences are helpful for pursuing the instrumentally desirable, he is talking about *certain* moral sentences; those pertaining, I take it, to a certain moral system. What he claims, then, is that there is one specific moral discourse that should be preserved because of its instrumental benefits. The fictionalist's promise that her theory will not imply revisions, therefore, will be kept only for some: for those, of course, who already use the discourse in question. Let us assume, though, that this exclusive sort of anti-revisionism is good enough.

If all of the above is right, Joyce's proposal and Kalderon's are different in a very important respect. Kalderon holds, we have seen, a pluralistic-relativistic view of morality: he maintains that there is no ultimate moral truth, but there exist a plurality of moral fictions that are all legitimate—and equally so. On Joyce's view, by contrast, there is one moral fiction that is superior to the others: one fiction that helps us attain that which becomes desirable once we realise that there are no moral facts.

In the previous section, I showed that Kalderon's pluralism-relativism is at odds with the fictionalist's anti-revisionist aim: if there are multiple moral practices that are equally legitimate, then every principle that permits or recommends interventionism or intolerance ought to be revised; if there is no moral truth, then we may just as well part from morality completely, and thereby perform the most radical of moral revisions. Joyce's view is not like this at all. Joyce contends that there are powerful reasons not to abandon morality, and he denies that all moral practices stand on equal footing.

If this is the case, I believe, then it appears that Joyce's theory lacks precisely that feature that led Kalderon astray from his anti-revisionist aim. It appears, therefore, that Joyce's theory is truly promising with regard to anti-revisionism. In the following section, I will investigate whether it is only promising, or also effectively successful.

4 Revisionism and Eliminativism

Joyce, we saw earlier, maintains that there are three alternative ways to act in response to the moral error: eliminativism, self-deception, and fictionalism. He also maintains that fictionalism—as formulated by him, at least—is superior to both of its alternatives. I think that he is wrong, however.

Joyce claims that fictionalism is not a form of self-deception. It seems to me, though, that this claim is untenable. Joyce's proposal is that, in order to outpower *akrasia*, the principles of our actions must be those that possess the special must-be-doneness that characterise the imperatives of morality. On Joyce's view, then, the imperatives that we must *act on* are not instrumental imperatives, but the categorical imperatives of morality. For fictionalists, however, morality is a fiction—and, to my mind, acting on merely fictional imperatives is clearly to be self-deceived. If this is right, then fictionalism is, *pace* Joyce, not preferable to the second possible response to the moral error.

Joyce, we have seen, believes that there is a moral system whose ends coincide with the instrumentally desirable. He claims, moreover, that moral discourse may be rationally preserved on the face of moral error. Since he addresses this claim to his readers, I will assume, for the time being, that the moral system that Joyce has in mind is one that his readers would very plausibly accept. In what follows, I will assume very little about this system: I will assume only that it forbids genocide, slavery and abuse. Accordingly, I will henceforth use the term 'morality' to denote this very minimal system.

To my mind, Joyce's claim is widely implausible that morality (as I just defined it) coincides with the instrumentally desirable. It is evident that there are cases where morality is radically opposed to that which is instrumentally desirable. Suppose, for instance, that a is terminally ill, and wishes intensely to perform an action φ that will cause serious harm to another human being. a has no reason to fear the consequences that her actions might accrue to her: she is with all certainty going to die very soon, so there will be no time for her left to suffer any such consequences. Clearly, morality

dictates that a does not φ. Nevertheless, it seems unquestionable that what is instrumentally best for a in these circumstances is to φ, for a desires to φ, and φ will come at not cost to her at all. In this case, then, morality and the instrumentally desirable do not coincide.

Now, Joyce seems to suggest at one point that morality leads to the instrumentally desirable not in every single case, but only "generally" (2001, p. 184). Accordingly, he suggests that the fictionalist theory that he recommends allows us not to preserve our current moral discourse in its entirety, but to preserve significant portions of it (2001, p. 185). Let us suppose that establishing this partial form of non-revisionism is good enough for a fictionalist to attain the fictionalist goal that I described above. Unfortunately, the one argument that Joyce (2001, pp. 181–185) advances in order to motivate his claim that "correct moral thinking" generally leads to the instrumentally desirable is, it seems to me, deeply flawed. Joyce discusses the case of a powerful empire whose leaders discover a truly desirable piece of land inhabited by indigenous peoples. Morality, Joyce claims, dictates that the leaders of the powerful empire should not enslave or exterminate the indigenous peoples, but, instead, foster a cooperative relation with them. And cooperation, Joyce contends, is also the instrumentally best course of action: cooperation, Joyce suggests, would in itself bring greater instrumental benefits to the empire; and, moreover, cooperating, rather than exterminating and enslaving, would allow the empire to preserve a good reputation on the face of other societies. To my mind, though, there are several independent arguments that demonstrate that Joyce's contentions are all false.

For one thing, nothing guarantees that cooperation is even possible between the indigenous peoples and the empire. If there is no end that both parties are interested in, then cooperation is flat-out impossible. Importantly, moreover, there might not be any one such end: perhaps the indigenous peoples wish only to be left alone. Moreover, regardless of what ends are those that the empire pursues, it is clear that the most instrumentally efficient way that there is for it to pursue its ends consists in enslaving the indigenous peoples so that they forcibly work, on behalf of the empire, towards the ends in question. If the empire seeks simply that which is instrumentally best, why would it ever waste resources on helping the indigenous peoples attain the indigenous people's ends, if it might instead use the indigenous peoples as tools in its own pursuits? Joyce seems to suggest an answer to this question: if the empire relates to these indigenous peoples in a hostile manner, it will damage its reputation. It seems to me, however, that

199

this answer makes no sense within Joyce's own fictionalist framework. If we all abandon morality, as Joyce recommends that we do, why would a third party, in the scenario that we are considering, judge the empire negatively for enslaving the indigenous peoples? If morality is an error, and the only thing that an agent may rationally pursue is the instrumentally desirable, then enslaving the indigenous peoples is, for the empire, nothing other than the rational thing to do. There is, then, no reason why a third party should judge the empire negatively. Perhaps, however, Joyce's thought is that third parties may become mistrustful enough of the empire to ward against cooperation with it—but then again the question arises of why the empire would wish to cooperate with third parties if it can exterminate them or enslave them instead.

If my arguments above are sound, then Joyce has no sound case in favour of his contention that morality coincides the instrumentally desirable. Recall, though, that I have been using 'morality' to refer to a particular moral system—a minimal system that Joyce's readers very plausibly endorse. Perhaps, however, there is a different moral system that does permit the genocide and the enslavement of peoples, and, therefore, a moral system that does coincide with the instrumentally desirable. Let us call this system S. Perhaps it is the advocates of S that Joyce's anti-revisionist contention is true for. If the categorical imperatives of S lead to the instrumentally desirable, then the advocates of S will have a reason to preserve *their* moral discourse after discovering that morality is an error. Nevertheless, Joyce has not showed that S actually exists; he has not shown us that there really are advocates of a system like S. Since he has not done so, we have no reason whatsoever to believe that persons exist whose moral commitments coincide, in some sort of generality, with the instrumentally desirable. Consequently, we have, at present, no reason to believe that there are persons who can rationally preserve moral discourse on the face of the moral error.

This conclusion produces two important results. The first is that Joyce has failed to prove that his theory fosters an anti-revisionary attitude towards moral discourse. Joyce has failed, then, like Kalderon, in attaining the supreme fictionalist goal of encompassing anti-realism and anti-revisionism. The second result is that fictionalism is not, after all, preferable to eliminativism. I argued that the moral system that we currently endorse does not, in any sort of generality, promote the instrumentally desirable. Our current categorical imperatives, therefore, will not lead us to any instrumentally desirable ends. Nevertheless, Joyce believes that the only thing that may rationally pursue on the face of moral error is the instrumentally desirable. If my arguments are sound and Joyce is

right, then, on the face of moral error, we ought to pursue the instrumentally desirable, and we ought to do so without the aid of the categorical imperatives of our moral system. But that is exactly what the eliminativist does: one the face of moral error, she abandons moral discourse and proceeds to pursue the instrumentally desirable with the resources that she has left.

From the above it follows that fictionalism is not preferable to eliminativism. Importantly, I argued earlier that fictionalism is not preferable to the self-deception response to moral error. If my arguments are sound, then Joyce has not only failed to attain the supreme fictionalist goal: he has also failed to give us a reason to prefer his theory over its non-fictionalist alternatives. If this is the case, then Joyce's fictionalism cannot be rationally endorsed.

Perhaps advocates are, sadly, not too hard to find of a moral system like S—a system, that is to say, that permits the genocide and the enslavement of peoples. If Joyce showed us that such advocates exist, and told us that it is them for whom his anti-revisionary promise is kept, then, I believe, the following question would present itself to us: why should *we* accept Joyce's fictionalist theory? Did we not want a fictionalist theory to endorse moral anti-realism without paying the price of having to revise our use of moral discourse? If Joyce's theory does not allow *us* to escape they payment of said price, there is simply to reason for *us* to adhere to it. The existence of the advocates of S, in sum, would force me to weaken my contentions, but not in any way that is relevant for us: if the advocates of S exist, then it is rational for them to endorse Joyce's fictionalism; for us, it remains true that fictionalism cannot be rationally endorsed.

5   Conclusion

In §2, I argued that Kalderon's fictionalist theory of morality promotes moral revisionism. It thus fails, I contended, to comply with the supreme fictionalist goal of encompassing anti-realism and anti-revisionism. In §3, I presented Joyce's version of moral fictionalism, and I showed that it lacks precisely that feature that led Kalderon astray from his anti-revisionist aim. I suggested, accordingly, that, if a fictionalist theory of morality has good chances of succeeding in the fictionalist's pursuit for anti-revisionism, it is Joyce's. Nevertheless, in §4 I argued that Joyce has in fact failed to prove that his theory fosters an anti-revisionary attitude towards moral discourse.

Additionally, in §1 I argued, contra Kalderon, that his theory of morality is not supe-

rior to its traditional, non-factualist rival. In §4 I contended, similarly, that Joyce's fictionalism is not preferable in any way to the other possible responses that he himself countenances to the discovery of the moral error. I thus conclude that fictionalists have not only failed in attaining the supreme fictionalist goal: they have also failed to give us a reason to prefer their theory over its alternatives. If this is the case, then fictionalist cannot be rationally endorsed.

This is the end of Part Two and, hence, I wish to present a brief conclusion concerning the entirety of the part in question. In Chapter Seven, I discussed the nominalistic impurity objection against mathematical fictionalism, and I argued that this objection entails that mathematical fictionalists cannot endorse an anti-revisionary view of mathematical statements—not, at least, without renouncing their nominalist commitments, or attributing to these statements a non-standard semantics. In Chapter Nine, I argued that both Joyce's and Kalderon's versions of moral fictionalism fail to be anti-revisionary. It follows, then, that both mathematical and modal fictionalism may be objected on the same grounds: their advocates face serious difficulties in sustaining their anti-revisionary convictions.

Now, modal fictionalists, we have seen, are anti-realists about possible worlds: they believe, that is to say, that possible worlds do not exist. Accordingly, they maintain that the theory that says that possible worlds exist is a mere fiction. On their view, moreover, modal statements express propositions that concern the contents of this fiction. In Chapter Eight, I argued that modal fictionalists have not yet succeeded in answering the following question satisfactorily: amongst the wealth of fictions that exist in Platonic heaven about alternative universes, which one is the one that modal statements concern? If my arguments are sound, then modal fictionalists have failed to tell us what, on their view, modal statements are about. Consequently, they have failed in demonstrating that modal statements express propositions that we may rationally assert even if we reject the existence of possible worlds. If this is the case, then modal fictionalists have failed in proving that an anti-revisionary stance towards modal statements can be rationally upheld within their framework of possible-world anti-realism.

If all of the above is correct, then not only two, but all three of the fictionalist theories that I discussed in Part Two can be objected on the grounds that their advocates face serious difficulties in sustaining their anti-revisionary convictions. To my mind, this strongly suggests that fictionalist theories in general might be flawed in a system-

atic manner. Investigating whether this hypothesis is ultimately right or not, I think, constitutes a potentially interesting path of inquiry. Independently of whether or not the hypothesis in question is right, however, I believe that I have, throughout my thesis, provided robust support to the contention that fictionalist theories in general are severely defective. In Part One, I argued (amongst other things) that fictionalists about all spheres of discourse have thus far failed in blocking the failure result that I described in the Introduction. According to this result, fictionalists are ontologically committed to exotic entities—to the entities, that is to say, that they were set to reject. In Part Two, I argued that fictionalists about three distinct spheres of discourse have failed to uphold their anti-revisionary claims. If my arguments are sound, then fictionalist theories in general are unable to offer to us a naturalistic ontology, and it is unclear whether any one of them can offer to us a non-revisionary view of our discourses. Since fictionalists in general aimed at construing semantically standard, ontologically naturalistic, and anti-revisionary theories of our discourses, then the outcomes of their efforts are, I conclude, severely defective.

Bibliography

Adams, R. M. (1974). Theories of Actuality. *Noûs*, 8, pp. 211-–31.

Adams, R. M. (1979). Primitive Thisness and Primitive Identity. *Journal of Philosophy* 76 (1), pp. 5–26.

Aliyev, A. (2020). The Proportionality Argument and the Problem of Widespread Causal Overdetermination. *Disputatio* 12 (59), pp. 331–355.

Aliyev, A. (forthcoming). Paraphrasability and the Causal Status of Types. *Theoria.*

Assadian, B. (2018). The Semantic Plights of the Ante-Rem Structuralist. *Philosophical Studies* 175 (12), pp. 1–20.

Assadian, B. (2019). In Defence of Utterly Indiscernible Entities. *Philosophical Studies* 176 (10), pp. 2551–2561.

Armstrong, D. M. (1989). *A Combinatorial Theory of Possibility.* Cambridge University Press.

Bach, K. (1987). *Thought and Reference.* Oxford University Press.

Balaguer, M. (1998). *Platonism and Anti-Platonism in Mathematics.* Oxford University Press.

Benacerraf, P. (1965). What Numbers Could Not Be. *Philosophical Review* 74 (1), pp. 47–73.

Bennett, J. (1988). *Events and Their Names.* Cambridge University Press.

Black, M. (1952). The Identity of Indiscernibles. "Mind" 61 (242), pp. 153-164.

Borges, J. L. (1944a). Funes, el Memorioso. In *Ficciones.* Alianza.

Borges, J. L. (1944b). Pierre Menard, Autor del Quijote. In *Ficciones.* Alianza.

Braun, D. (1993). Empty Names. *Noûs* 27 (4), pp. 449–469.

Braun, D. (2005). Empty Names, Fictional Names, Mythical Names. *Noûs* 39 (4), pp. 596–631.

Breckenridge, W. & Magidor, O. (2012). Arbitrary Reference. *Philosophical Studies* 158 (3), pp. 377–400.

Brock, S. (1993). Modal Fictionalism: A Response to Rosen. *Mind* 102(405), pp. 147-–150.

Brock, S, (2002). Fictionalism about Fictional Characters. *Noûs* 36 (1), pp. 1–21.

Brock, S. (2016). Fictionalism about Fictional Characters Revisited. *Res Philosophica* 93 (2), pp. 377–403.

Burgess, J. & Rosen, G. (1997). *A Subject With No Object.* Oxford: Clarendon Press.

Butterfield, J. & Caulton, A. (2012). On Kinds of Indiscernibility in Logic and Metaphysics. *British Journal for the Philosophy of Science* 63 (1), pp. 27–84.

Button, T. (2012). Spotty Scope and Our Relation to Fictions. *Noûs* 46 (2), pp. 243–58.

Button, T. & Walsh, S. (2016). Structure and Categoricity: Determinacy of Reference and Truth-Value in the Philosophy of Mathematics. *Philosophia Mathematica*, 24(3), pp. 283–307.

Caplan, B. (2016). The Extraordinary Impossibility of Sherlock Holmes. *Res Philosophica* 93 (2): 335-355.

Caplan, B. and Matheson, C. (2004). Can a Musical Work Be Created? *British Journal of Aesthetics* 44, pp. 113-–134.

Caplan, B. & Matheson, C. (2006). Defending Musical Perdurantism. *British Journal of Aesthetics* 46 (1), pp. 59–69.

Chihara, C. (1990), *Constructibility and Mathematical Existence.* Oxford University Press.

Child, W. (2019). Meaning, Use, and Supervenience. In J. Conant and S. S. Grève (eds.), *Wittgenstein on Philosophy, Objectivity, and Meaning.* Cambridge University Press.

Cohen, W. A. (forthcoming). The Audibility Problem and Indirect Listening. *Australasian Journal of Philosophy.*

Cortázar, J. (1962). *Historias de Cronopios y Famas.* Minotauro.

Crimmins, M. (1998). Hesperus and Phosphorus: Sense, Pretense, and Reference. *Philosophical Review* 107 (1), pp. 1–47.

Daly, C. J. (2008). Fictionalism and the Attitudes. *Philosophical Studies* 139 (3), pp. 423–440.

Daly, C. & Liggins, D. (2011). Deferentialism. *Philosophical Studies* 156, pp. 321-337.

Davidson, D. (1963). Actions, Reasons and Causes. *Journal of Philosophy* 60.

Davidson, D. (1967). Causal Relations. *Journal of Philosophy* 64.

Davies, D. (2004). *Art as Performance.* Blackwell.

Davies, D. (2009). Dodd on the Audibility of Musical Works. *British Journal of Aesthetics* 49 (2), pp. 99–108.

Divers, J. (1999). A Modal Fictionalist Result. *Noûs* 33(3), pp. 317—346.

Dodd, J. (2000). Musical Works as Eternal Types. *British Journal of Aesthetics* 40 (4), pp. 424–440.

Dodd, J. (2002). Defending Musical Platonism. *British Journal of Aesthetics* 42 (4), pp. 380–402.

Dodd, J. (2004). Types, Continuants, and the Ontology of Music. *British Journal of Aesthetics* 44 (4), pp. 342-360.

Dodd, J. (2007). *Works of Music: an Essay in Ontology.* Oxford University Press.

Dodd, J. (2008). Musical works: Ontology and Meta-Ontology. *Philosophy Compass* 3 (6), pp. 1113–1134.

Dodd, J. and Letts, P. (2017). Types, Tokens, and Talk about Musical Works. *Journal of Aesthetics and Art Criticism* 75 (3), pp. 249–263.

Dorsch, F. (2013). Non-Inferentialism about Justification. *Philosophical Quarterly* 63 (253), pp. 660–682.

Dummett, M. (1973). *Frege: Philosophy of language.* Harper and Row.

Eklund, M. (2009). The Frege-Geach Problem and Kalderon's Moral Fictionalism. *Philosophical Quarterly* 59 (237), pp. 705–712.

Eklund, M. (2019). Fictionalism. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy.*

Evans, G. (1982). *The Varieties of Reference.* Clarendon Press.

Everett, A. (2005). Against Fictional Realism. *Journal of Philosophy* 102 (12), pp. 624–649.

Everett, A. (2013). *The Nonexistent.* Oxford University Press.

Falguera, J. L., Martínez-Vidal, C., & Rosen, G. (2022). Abstract Objects. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy.*

Feldman, F. (1971). Counterparts. *Journal of Philosophy* 68 (13), pp. 406–409.

Field, H. (1980). *Science Without Numbers.* Oxford University Press.

Field, H. (1989). *Realism, Mathematics and Modality.* Blackwell.

Fine, K. (1985a). Natural Deduction and Arbitrary Objects. *Journal of Philosophical Logic* 14 (1), pp. 57–107.

Fine, K. (1985b). *Reasoning with Arbitrary Objects.* Basil Blackwell.

Friedell, D. (2020a). Abstracta Are Causal. *Philosophia* 48 (1), pp. 33–142.

Friedell, D. (2020b). Fiction and Indeterminate Identity. *Analysis* 80 (2), pp. 221–229.

Forrest, P. (2020). The Identity of Indiscernibles. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy.*

Friend, S. (2011). The Great Beetle Debate: a Study in Imagining with Names. *Philosophical Studies* 153 (2), pp. 183–211.

Friend, S. (2014). Notions of Nothing. In M. García-Carpintero and G. Martí (eds.), *Empty Representations: Reference and Non-Existence.* Oxford University Press.

Gasparri, L. (2021). Semantic Supervenience. *Inquiry: An Interdisciplinary Journal of Philosophy.*

Geach, P. (1965). Assertion. *Philosophical Review* 74, pp. 449—465.

Goodman, N. (1977). *Structure of Appearance.* Reidel.

Goodman, N. & Quine W. V. O. (1947). Steps Toward a Constructive Nominalism. *Journal of Symbolic Logic* 12, pp. 105-–122.

Hacking, I. (1975). The Identity of Indiscernibles. *Journal of Philosophy* 72 (9), pp. 249-–256.

Hale, B. (1995). Modal Fictionalism: A Simple Dilemma. *Analysis* 55(2), pp. 63-–67

Haraldsen, F. (2017). The Truth About Sherlock Holmes. *Organon F* 24 (3), pp. 339–365.

Haraldsen, F. (2020). How to Pick Out a Dragon: Fiction and the Selection Problem. *Topoi* 39 (2), pp. 401–412.

Haze, T. (2016). A Counterexample to the Breckenridge-Magidor Account of Instantial Reasoning. *Journal of Philosophical Research* 41, pp. 539–541.

Hellman, G. (1996). Structuralism Without Structures. *Philosophia Mathematica*, 4(2), pp. 100—123.

Hellman, G. (2001). Three Varieties of Mathematical Structuralism. *Philosophia Mathematica* (3) 9, pp. 184--211.

Hellman, G. (2005). Structuralism. In S. Shapiro (ed.), *The Oxford Handbook of Philosophy of Mathematics and Logic.* Oxford University Press.

Howell, R. (2002). Types, Indicated and Initiated. *British Journal of Aesthetics* 42 (2), pp. 105–127.

Jacquette, D. (1996). *Meinongian logic: the Semantics of Existence and Nonexistence.* De Gruyter.

Joyce, R. (2001). *The Myth of Morality.* Cambridge University Press.

Juvshik, T. (2018). Abstract Objects, Causal Efficacy, and Causal Exclusion. *Erkenntnis* 83 (4), pp. 805–827.

Juvshik, T. (2020). Relativity and the Causal Efficacy of Abstract Objects. *American Philosophical Quarterly* 57 (3).

Kalderon, M. E. (2005). *Moral Fictionalism.* Clarendon Press.

Kearns, S. & Magidor, O. (2012). Semantic Sovereignty. *Philosophy and Phenomenological Research* 82 (2), pp. 322–350.

Khlebnikov, V. (2000). *Sobranie Sochinenii v Shesti Tomakh.* IMLI RAN.

Kim, J. (1986). Possible Worlds and Annstrong's Combinatorialism. *Canadian Journal of Philosophy* 16 (4), pp. 595–612.

Kim, J. (1993). *Supervenience and Mind.* Cambridge University Press.

Kim, J. (1998). *Mind in a Physical World: an Essay on the Mind-Body Problem and Mental Causation.* MIT Press.

King, J. C. (1991). Instantial Terms, Anaphora and Arbitrary Objects. *Philosophical Studies* 61 (3), pp. 239–265.

Kivy, P. (1983). Platonism in Music: A Kind of Defense. *Grazer Philosophische Studien* 19 (1), pp. 109–129.

Kivy, P. (1987). Platonism in Music: Another Kind of Defense. *American Philosophical Quarterly* 24 (3), pp. 245–252.

Kivy, P. (2002). *Introduction to a Philosophy of Music.* Clarendon Press.

Kripke, S. (1977). Speaker's Reference and Semantic Reference. *Midwest Studies in Philosophy* 2 (1), pp. 255–276.

Kripke, S. (1980). *Naming and Necessity.* Harvard University Press.

Kripke, S. A. (2013). *Reference and Existence: the John Locke Lectures for 1973.* Oxford University Press.

Kroon, F. (2000). "Disavowal Through Commitment" Theories of Negative Existentials. In A. Everett and T. Hofweber (eds.), *Empty Names, Fiction, and the Puzzles of Non-Existence.* CSLI Press.

Kroon, F. (2004). Descriptivism, Pretense, and the Frege–Russell Problems. *Philosophical Review* (57), pp. 1-–30.

Kroon, F. & Voltolini, A. (2022). Fictional Entities. In E. N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy.*

Lebens, S. (2015). Would This Paper exist if I Hadn't Written It? *Philosophical Studies* 172 (11), pp. 3059–3080.

Leitgeb, H. (2022). On Non-Eliminative Structuralism. Unlabeled Graphs as a Case Study, Part B. *Philosophia Mathematica.*

Leng, M. (2010). *Mathematics and Reality.* Oxford University Press.

Lenman, J. (2008). Against Moral Fictionalism. *Philosophical Books* 49 (1), pp. 23–32.

Leslie, S. J. & Lerner, A. (2022). Generic Generalizations. In E. N. Zalta & U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy.*

Levinson, J. (1980). What a Musical Work Is. *Journal of Philosophy* 77 (1), pp. 5–28.

Lewis, D. (1971). Counterparts of Persons and Their Bodies. *Journal of Philosophy* 68 (7), pp. 203–211.

Lewis, D. K. (1973). *Counterfactuals.* Blackwell.

Lewis, D. K. (1975). Adverbs of Quantification. In E. L. Keenan (ed.), *Formal Semantics of Natural Language.* Cambridge University Press.

Lewis, D. K. (1978). Truth in Fiction. *American Philosophical Quarterly* 15 (1), pp. 37–46.

Lewis, D. (1983). Individuation by Acquaintance and by Stipulation. *Philosophical Review* 92 (1), pp. 3–32.

Lewis, D. (1986). *On The Plurality of Worlds.* Blackwell.

Liao, S. & Gendler, T. (2020). Imagination. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy.*

Linsky, B. & Zalta, E. (1995). Naturalized Platonism Versus Platonized Naturalism. *The Journal of Philosophy* 92, pp. 525-–555.

Maddy, P. (1990). *Realism in Mathematics.* Oxford University Press.

Malament, D. (1982), Review of Field, Science Without Numbers. *Journal of Philosophy.* 79, pp. 523-–34.

McDaniel, K. (2004). Modal Realism with Overlap. *Australasian Journal of Philosophy* 82, pp. 137-–152.

McGee, V. (1997). How We Learn Mathematical Language. *Philosophical Review* 106, pp. 35-–68.

Meléndez Gutiérrez, S. (forthcoming a). Musical Ontology and the Audibility of Musical Works. *British Journal of Aesthetics.*

Meléndez Gutiérrez, S. (forthcoming b). A Reply to Haze's Argument Against Arbitrary Reference. *Philosophia.*

Melia, J. (2000). Weaseling away the Indispensability Argument. *Mind* 109 (435), pp. 455–480.

Menzel, C. (2021) Possible Worlds. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy.*

Menzies, P. and Pettit, P. (1994). In Defence of Fictionalism about Possible Worlds. *Analysis* 54(1), pp. 27-–36.

Moruzzi, C. (2018). Every Performance Is a Stage: Musical Stage Theory as a Novel Account for the Ontology of Musical Works. *Journal of Aesthetics and Art Criticism* 76 (3), pp. 341–351.

Nolan, D. (1997). Three Problems for 'Strong' Modal Fictionalism. *Philosophical Studies* 87(3), pp. 259-–275.

Nolan, D. (2022). Modal Fictionalism. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy.*

Noonan, H. (1994). In Defence of the Letter of Fictionalism. *Analysis* 54(3), pp. 133—139.

Orilia, F. & Paolini Paoletti, M. (2022). Properties. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy.*

Ostertag, G. 2012. Critical Study: Julian Dodd. Works of Music: An Essay in Ontology. *Noûs* 46, pp. 355–374.

Parsons, T. (1980). *Nonexistent Objects.* Yale University Press.

Parsons, C. (2008). *Mathematical Thought and its Objects.* Cambridge University Press.

Peacocke, C. (1999). *Being Known.* Oxford University Press.

Pratt, T. (1998). From the Andes to Epcot: the Adventures of an 8,000-Year-Old-Bean. *The New York Times*, May 18, 1998.

Predelli, S. 2011. Talk about Music: From Wolterstorffian Ambiguity to Generics. *The Journal of Aesthetics and Art Criticism* 69, pp. 273–283.

Puy, N. G. C. (2019). The Ontology of Musical Versions: Introducing the Hypothesis of Nested Types. *Journal of Aesthetics and Art Criticism* 77 (3), pp. 241–254.

Puy, N. G. C. (2021). Julian Dodd. Being True to Works of Music. *Journal of Aesthetics and Art Criticism* 79 (2), pp. 268–272.

Quine, W. V. O. (1960). *Word and Object.* MIT Press.

Quine, W. V. O. (1961). *From a Logical Point of View.* Harper and Row.

Quine, W.V.O. (1969). *Ontological Relativity, in Ontological Relativity and Other Essays.* Columbia University Press.

Resnik, M. (1985). How Nominalist is Hartry Field's Nominalism? *The Philosophical Review* 117, pp. 385-–443.

Resnik, D. M. (1997). *Mathematics as a Science of Patterns.* Oxford University Press.

Riffaterre, M. (1990). *Fictional Truth.* Johns Hopkins University Press.

Rosen, G. (1990). Modal Fictionalism. *Mind* 99 (395), pp. 327–354.

Rosen, G. (1993). A Problem for Fictionalism About Possible Worlds. *Analysis* 53(2), pp. 71–81.

Rosen, G. (1995). Modal Fictionalism Fixed. *Analysis* 55(2), pp. 67-–73.

Sainsbury, R. M. (2009). *Fiction and Fictionalism.* Routledge.

Salmon, N. (1998). Nonexistence. *Noûs* 32 (3), pp. 277–319.

Searle, J. (1962). Meaning and Speech Acts. *Philosophical Review*, 71, pp. 423-–32.

Sellars, W. (1963). Abstract Entities. *Review of Metaphysics* 16, pp. 627-–671.

Sennet, A. (2021). Ambiguity. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy.*

Shapiro, S. (1983). Conservativeness and Incompleteness. *Journal of Philosophy* 80, pp. 521-–31.

Shapiro, S. (1997). *Philosophy of Mathematics: Structure and Ontology.* Oxford University Press.

Shapiro, S. (2012). An '*i*' for an *i*: Singular Terms, Uniqueness, and Reference. *Review of Symbolic Logic* 5 (3), pp. 380–415.

Skiba, L. (2017). Fictionalism and the Incompleteness Problem. *Synthese* 194, pp. 1349-–1362.

Stanley, J. (2001). Hermeneutic Fictionalism. In P. French and H. Wettstein, (eds.), *Midwest Studies in Philosophy*, Volume XXV. Blackwell.

Stokke, A. (2020). Fictional Names and Individual Concepts. *Synthese* 198 (8), pp. 7829–7859.

Stone, J. (2010). Harry Potter and the Spectre of Imprecision. *Analysis* 70 (4), pp. 638–644.

Thomasson, A. L. (1998). *Fiction and Metaphysics.* Cambridge University Press.

Updike, T. (1991). The Other Side of the Street. *The New Yorker*, October 28, 1991.

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

Van Inwagen, P. (1983). Fiction and Metaphysics. *Philosophy and Literature* 7 (1), pp. 67–77.

Van Inwagen, P. (2000). Quantification and Fictional Discourse. In A. Everett and T. Hofweber (eds.), *Empty Names, Fiction and the Puzzles of Non-Existence.* CSLI Publications.

Van Inwagen, P. (2007). A Materialist Ontology of the Human Person. In Van P. Inwagen and D. Zimmerman (eds.), *Persons: Human and Divine.* Oxford University Press.

Voltolini, A. (2006). *How Ficta Follow Fiction.* Springer.

Von Solodkoff, T. (2014a). Fictional Realism and Negative Existentials. In M. García-Carpintero and G. Martí (eds.), *Empty Representations: Reference and Non-Existence.* Oxford University Press.

Von Solodkoff, T. (2014b). Paraphrase Strategies in Metaphysics. *Philosophy Compass* 9 (8), pp. 570–582.

Von Solodkoff, T. and Woodward, R. (2017). To Have and to Hold. *Philosophical Issues* 27 (1), pp. 407–427.

Walters, L. (2013). Repeatable Artworks as Created Types. *British Journal of Aesthetics* 53 (4), pp. 461–477.

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

Walton, K. (2000). Existence as Metaphor? In A. Everett and T. Hofweber (eds.), *Empty Names, Fiction, and the Puzzles on Non-Existence.* CSLI Press.

Wellek, R., & Warren, A. (1956). *Theory of Literature.* Harcourt Brace & Company.

Wetzel, L. (2009). Types and Tokens: on Abstract Objects. *MIT Press.*

Wetzel, L. (2018). Types and Tokens. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy.*

Wollheim, R. (1980). *Art and its Objects.* Cambridge University Press.

Wolterstorff, N. (1980). *Works and Worlds of Art.* Oxford University Press.

Woodward, R (2011). Is Modal Fictionalism Artificial? *Pacific Philosophical Quarterly* 92 (4), pp. 535–550.

Woodward, R. (2012). Fictionalism and Incompleteness. *Noûs* 46(4), pp. 781-–790.

Woodward, R. (2017). Identity in Fiction. *Philosophy and Phenomenological Research* 94, pp. 646-–71.

Yablo, S. (2001). Go Figure: a Path through Fictionalism. *Midwest Studies in Philosophy* 25 (1), pp. 72-–102.

Yablo, S. (2005). The Myth of Seven. In M. E. Kalderon (ed.), *Fictionalism in Metaphysics.* Clarendon Press, pp. 88–115.

Yablo, S. (2012). Explanation, Extrapolation, and Existence. *Mind* 121 (484), pp. 1007–1029.

Yagisawa, T. (2010). *Worlds and Individuals, Possible and Otherwise.* Oxford University Press.

Zalta, E. N. (1983). *Abstract Objects: An Introduction to Axiomatic Metaphysics.* D. Reidel.

Zouhar, M. (2019). On the Systematic Inadequacy of Fictionalism about Fictional Characters. *Philosophia* 47 (3), pp. 925–942.

Zwicky, A. and Sadock, J. (1975) 'Ambiguity Tests and How to Fail Them'. In J. P. Kimball (ed.), *Syntax and Semantics 4*, pp. 1–35.